Technical Report # KU-EC-11-1: Distribution of the Relative Density of Central Similarity Proximity Catch Digraphs Based on One Dimensional Uniform Data

Elvan Ceyhan*

January 26, 2011

short title: Relative Density of Central Similarity Proximity Catch Digraphs

Abstract

We consider the distribution of a graph invariant of central similarity proximity catch digraphs (PCDs) based on one dimensional data. The central similarity PCDs are also a special type of parameterized random digraph family defined with two parameters, a centrality parameter and an expansion parameter, and for one dimensional data, central similarity PCDs can also be viewed as a type of interval catch digraphs. The graph invariant we consider is the relative density of central similarity PCDs. We prove that relative density of central similarity PCDs is a U-statistic and obtain the asymptotic normality under mild regularity conditions using the central limit theory of U-statistics. For one dimensional uniform data, we provide the asymptotic distribution of the relative density of the central similarity PCDs for the entire ranges of centrality and expansion parameters. Consequently, we determine the optimal parameter values at which the rate of convergence (to normality) is fastest. We also provide the connection with class cover catch digraphs and the extension of central similarity PCDs to higher dimensions.

Keywords: asymptotic normality; class cover catch digraph; intersection digraph; interval catch digraph; random geometric graph; U-statistics

AMS 2000 Subject Classification: 05C80; 05C20; 60D05; 60C05; 62E20

1 Introduction

Proximity catch digraphs (PCDs) are introduced recently and have applications in spatial data analysis and statistical pattern classification. The PCDs are a special type of proximity graphs which were introduced by Toussaint (1980). Furthermore, the PCDs are closely related to the class cover problem of Cannon and Cowen (2000). The PCDs are vertex-random digraphs in which each vertex corresponds to a data point, and directed edges (i.e., arcs) are defined by some bivariate relation on the data using the regions based on these data points.

*Address: Department of Mathematics, Koç University, 34450 Sarıyer, Istanbul, Turkey. e-mail: elceyhan@ku.edu.tr, tel:+90 (212) 338-1845, fax: +90 (212) 338-1559.

Priebe et al. (2001) introduced the class cover catch digraphs (CCCDs) in $\mathbb{R}$ which is a special type of PCDs and gave the exact and the asymptotic distribution of the domination number of the CCCDs based on data from two classes, say $\mathcal{X}$ and $\mathcal{Y}$, with uniform distribution on a bounded interval in $\mathbb{R}$. DeVinney et al. (2002), Marchette and Priebe (2003), Priebe et al. (2003a), Priebe et al. (2003b), and DeVinney and Priebe (2006) applied the concept in higher dimensions and demonstrated relatively good performance of CCCDs in classification. Ceyhan and Priebe (2003) introduced central similarity PCDs for two dimensional data in an unparameterized fashion; the parameterized version of this PCD is later developed by Ceyhan et al. (2007) where the relative density of the PCD is calculated and used for testing bivariate spatial patterns in $\mathbb{R}^2$. Ceyhan and Priebe (2005, 2007), Ceyhan (2011b) applied the same concept (for a different PCD family called proportional-edge PCD) in testing spatial point patterns in $\mathbb{R}^2$. The distribution of the relative density of the proportional-edge PCDs for one dimensional uniform data is provided in Ceyhan (2011a).

In this article, we consider central similarity PCDs for one dimensional data. We derive the asymptotic distribution of a graph invariant called *relative (arc) density* of central similarity PCDs. Relative density is the ratio of number of arcs in a given digraph with n vertices to the total number of arcs possible (i.e., to the number of arcs in a complete symmetric digraph of order n). We prove that, properly scaled, the relative density of the central similarity PCDs is a U-statistic, which yields the asymptotic normality by the general central limit theory of U-statistics. Furthermore, we derive the explicit form of the asymptotic normal distribution of the relative density of the PCDs for uniform one dimensional $\mathcal{X}$ points whose support being partitioned by class $\mathcal{Y}$ points. We consider the entire ranges of the expansion and centrality parameters and the asymptotic distribution is derived as a function of these parameters based on detailed calculations. The relative density of central similarity PCDs is first investigated for uniform data in one interval (in $\mathbb{R}$) and the analysis is generalized to uniform data in multiple intervals. These results can be used in applying the relative density for testing spatial interaction between classes of one dimensional data. Moreover, the behavior of the relative density in the one dimensional case forms the foundation of our investigation and extension of the topic in higher dimensions.

We define the proximity catch digraphs and describe the central similarity PCDs in Section 2, define their relative density and provide preliminary results in Section 3, provide the distribution of the relative density for uniform data in one interval in Section 4 and in multiple intervals in Section 5, provide extension to higher dimensions in Section 6 and provide discussion and conclusions in Section 7. Shorter proofs are given in the main body of the article; while longer proofs are deferred to the Appendix Sections.

2 Vertex-Random Proximity Catch Digraphs

We first define vertex-random PCDs in a general setting. Let $(\Omega, \mathcal{M})$ be a measurable space and $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ and $\mathcal{Y}_m = \{Y_1, Y_2, \ldots, Y_m\}$ be two sets of Ω-valued random variables from classes $\mathcal{X}$ and $\mathcal{Y}$, respectively, with joint probability distribution $F_{X,Y}$ and marginals F_X and F_Y, respectively. A PCD is comprised by a set $\mathcal{V}$ of vertices and a set $\mathcal{A}$ of arcs. For example, in the two class case, with classes $\mathcal{X}$ and $\mathcal{Y}$, we choose the $\mathcal{X}$ points to be the vertices and put an arc from $X_i \in \mathcal{X}_n$ to $X_j \in \mathcal{X}_n$, based on a binary relation which measures the relative allocation of X_i and X_j with respect to $\mathcal{Y}$ points. Notice that the randomness is only on the vertices, hence the name *vertex-random PCDs*. Consider the map $N : \Omega \to \mathcal{P}(\Omega)$, where $\mathcal{P}(\Omega)$ represents the power set of Ω. Then given $\mathcal{Y}_m \subseteq \Omega$, the *proximity map* $N(\cdot)$ associates with each point $x \in \Omega$ a *proximity region* $N(x) \subseteq \Omega$. For $B \subseteq \Omega$, the Γ_1-region is the image of the map $\Gamma_1(\cdot, N) : \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$ that associates the region $\Gamma_1(B, N) := \{z \in \Omega : B \subseteq N(z)\}$ with the set B. For a point $x \in \Omega$, we denote $\Gamma_1(\{x\}, N)$ as $\Gamma_1(x, N)$. Notice that while the proximity region is defined for one point, a Γ_1-region is defined for a point or set of points. The *vertex-random PCD* has the vertex set $\mathcal{V} = \mathcal{X}_n$ and arc set $\mathcal{A}$ defined by $(X_i, X_j) \in \mathcal{A}$ if $X_j \in N(X_i)$. Let arc probability be defined as $p_a(i,j) := P((X_i, X_j) \in \mathcal{A})$ for all $i \neq j$, $i, j = 1, 2, \ldots, n$. Given $\mathcal{Y}_m = \{y_1, y_2, \ldots, y_m\}$, let $\mathcal{X}_n$ be a random sample from F_X. Then $N(X_i)$ are also iid and the same holds for $\Gamma_1(X_i, N)$. Hence $p_a(i,j) = p_a$ for all $i \neq j$, $i, j = 1, 2, \ldots, n$ for such $\mathcal{X}_n$.

2.1 Central Similarity PCDs for One Dimensional Data

In the special case of central similarity PCDs for one dimensional data, we have $\Omega = \mathbb{R}$. Let $Y_{(i)}$ be the i^{th} order statistic of $\mathcal{Y}_m$ for $i = 1, 2, \ldots, m$. Assume $Y_{(i)}$ values are distinct (which happens with probability one for continuous distributions). Then $Y_{(i)}$ values partition $\mathbb{R}$ into $(m+1)$ intervals. Let

$$-\infty =: Y_{(0)} < Y_{(1)} < \ldots < Y_{(m)} < Y_{(m+1)} := \infty.$$

We call intervals $(-\infty, Y_{(1)})$ and $\left(Y_{(m)}, \infty\right)$ the *end intervals*, and intervals $(Y_{(i-1)}, Y_{(i)})$ for $i = 2, \ldots, m$ the *middle intervals*. Then we define the central similarity PCD with the parameter $\tau > 0$ for two one dimensional data sets, $\mathcal{X}_n$ and $\mathcal{Y}_m$, from classes $\mathcal{X}$ and $\mathcal{Y}$, respectively, as follows. For $x \in \left(Y_{(i-1)}, Y_{(i)}\right)$ with $i \in \{2, \ldots, m\}$ (i.e., for x in a middle interval) and $M_c \in \left(Y_{(i-1)}, Y_{(i)}\right)$ such that $c \times 100$ % of $(Y_{(i)} - Y_{(i-1)})$ is to the left of M_c (i.e., $M_c = Y_{(i-1)} + c\,(Y_{(i)} - Y_{(i-1)})$)

$$N(x, \tau, c) = \begin{cases} \left(x - \tau\, \left(x - Y_{(i-1)}\right), x + \frac{\tau\,(1-c)\left(x - Y_{(i-1)}\right)}{c}\right) \bigcap \left(Y_{(i-1)}, Y_{(i)}\right) & \text{if } x \in (Y_{(i-1)}, M_c), \\ \left(x - \frac{c\,\tau\,\left(Y_{(i)} - x\right)}{1-c}, x + \tau\, \left(Y_{(i)} - x\right)\right) \bigcap \left(Y_{(i-1)}, Y_{(i)}\right) & \text{if } x \in \left(M_c, Y_{(i)}\right). \end{cases} \tag{1}$$

Observe that with $\tau \in (0, 1)$, we have

$$N(x, \tau, c) = \begin{cases} \left(x - \tau\, \left(x - Y_{(i-1)}\right), x + \frac{\tau\,(1-c)\left(x - Y_{(i-1)}\right)}{c}\right) & \text{if } x \in (Y_{(i-1)}, M_c), \\ \left(x - \frac{c\,\tau\,\left(Y_{(i)} - x\right)}{1-c}, x + \tau\, \left(Y_{(i)} - x\right)\right) & \text{if } x \in \left(M_c, Y_{(i)}\right), \end{cases} \tag{2}$$

and with $\tau \geq 1$, we have

$$N(x, \tau, c) = \begin{cases} \left(Y_{(i-1)}, x + \frac{\tau\,(1-c)\left(x - Y_{(i-1)}\right)}{c}\right) & \text{if } x \in \left(Y_{(i-1)}, \frac{c\,Y_{(i)} + \tau\,(1-c)\,Y_{(i-1)}}{c + \tau\,(1-c)}\right), \\ \left(Y_{(i-1)}, Y_{(i)}\right) & \text{if } x \in \left(\frac{c\,Y_{(i)} + \tau\,(1-c)\,Y_{(i-1)}}{c + \tau\,(1-c)}, \frac{(1-c)\,Y_{(i-1)} + c\,\tau\,Y_{(i)}}{1 - c + c\,\tau}\right), \\ \left(x - \frac{c\,\tau\,\left(Y_{(i)} - x\right)}{1-c}, Y_{(i)}\right) & \text{if } x \in \left(\frac{(1-c)\,Y_{(i-1)} + c\,\tau\,Y_{(i)}}{1 - c + c\,\tau}, Y_{(i)}\right). \end{cases} \tag{3}$$

For an illustration of $N(x, \tau, c)$ in the middle interval case, see Figure 1 (left) where $\mathcal{Y}_2 = \{y_1, y_2\}$ with $y_1 = 0$ and $y_2 = 1$ (hence $M_c = c$).

Additionally, for $x \in \left(Y_{(i-1)}, Y_{(i)}\right)$ with $i \in \{1, m+1\}$ (i.e., for x in an end interval), the central similarity proximity region only has an expansion parameter, but not a centrality parameter. Hence we let $N_e(x, \tau)$ be the central similarity proximity region for an x in an end interval. Then with $\tau \in (0, 1)$, we have

$$N_e(x, \tau) = \begin{cases} \left(x - \tau\, \left(Y_{(1)} - x\right), x + \tau\, \left(Y_{(1)} - x\right)\right) & \text{if } x < Y_{(1)}, \\ \left(x - \tau\, \left(x - Y_{(m)}\right), x + \tau\, \left(x - Y_{(m)}\right)\right) & \text{if } x > Y_{(m)} \end{cases} \tag{4}$$

and with $\tau \geq 1$, we have

$$N_e(x, \tau) = \begin{cases} \left(x - \tau\, \left(Y_{(1)} - x\right), Y_{(1)}\right) & \text{if } x < Y_{(1)}, \\ \left(Y_{(m)}, x + \tau\, \left(x - Y_{(m)}\right)\right) & \text{if } x > Y_{(m)}. \end{cases} \tag{5}$$

If $x \in \mathcal{Y}_m$, then we define $N(x, \tau, c) = \{x\}$ and $N_e(x, \tau) = \{x\}$ for all $\tau > 0$, and if $x = M_c$, then in Equation (1), we arbitrarily assign $N(x, \tau, c)$ to be one of $\left(x - \tau\, \left(x - Y_{(i-1)}\right), x + \frac{\tau\,(1-c)\left(x - Y_{(i-1)}\right)}{c}\right) \bigcap \left(Y_{(i-1)}, Y_{(i)}\right)$

or $\left(x - \frac{c\,\tau\,\left(Y_{(i)}-x\right)}{1-c}, x + \tau\,\left(Y_{(i)} - x\right)\right) \bigcap \left(Y_{(i-1)}, Y_{(i)}\right)$. For X from a continuous distribution, these special cases in the construction of central similarity proximity region — $X \in \mathcal{Y}_m$ and $X = M_c$ — happen with probability zero. Notice that $\tau > 0$ implies $x \in N(x,\tau,c)$ for all $x \in \left[Y_{(i-1)}, Y_{(i)}\right]$ with $i \in \{2,\dots,m\}$ and $x \in N_e(x,\tau)$ for all $x \in \left[Y_{(i-1)}, Y_{(i)}\right]$ with $i \in \{1, m+1\}$. Furthermore, $\lim_{\tau\to\infty} N(x,\tau,c) = \left(Y_{(i-1)}, Y_{(i)}\right)$ (and $\lim_{\tau\to\infty} N_e(x,\tau) = \left(Y_{(i-1)}, Y_{(i)}\right)$) for all $x \in \left(Y_{(i-1)}, Y_{(i)}\right)$ with $i \in \{2,\dots,m\}$ (and $i \in \{1, m+1\}$), so we define $N(x,\infty,c) = \left(Y_{(i-1)}, Y_{(i)}\right)$ (and $N_e(x,\infty) = \left(Y_{(i-1)}, Y_{(i)}\right)$) for all such x.

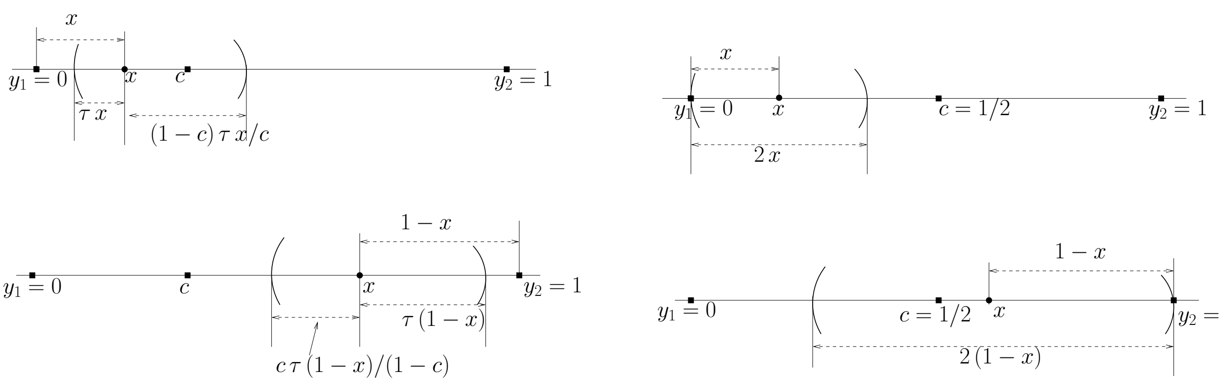

Figure 1: Plotted in the left is an illustration of the construction of central similarity proximity region, $N(x,\tau,c)$ with $\tau \in (0,1)$, $\mathcal{Y}_2 = \{y_1, y_2\}$ with $y_1 = 0$ and $y_2 = 1$ (hence $M_c = c$) and $x \in (0,c)$ (top) and $x \in (c,1)$ (bottom); and in the right is the proximity region associated with CCCD, i.e., $N(x,\tau=1,c=1/2)$ for an $x \in (0,1/2)$ (top) and $x \in (1/2,1)$ (bottom).

The vertex-random central similarity PCD has the vertex set $\mathcal{X}_n$ and arc set $\mathcal{A}$ defined by $(X_i, X_j) \in \mathcal{A} \iff X_j \in N(X_i,\tau,c)$ for X_i, X_j in the middle intervals and $(X_i, X_j) \in \mathcal{A} \iff X_j \in N_e(X_i,\tau)$ for X_i, X_j in the end intervals. We denote such digraphs as $\mathscr{D}_{n,m}(\tau,c)$. A $\mathscr{D}_{n,m}(\tau,c)$-digraph is a *pseudo digraph* according to some authors, if loops are allowed (see, e.g., Chartrand and Lesniak (1996)). The $\mathscr{D}_{n,m}(\tau,c)$-digraphs are closely related to the *proximity graphs* of Jaromczyk and Toussaint (1992) and might be considered as a special case of *covering sets* of Tuza (1994). Our vertex-random proximity digraph is not a standard random graph (see, e.g., Janson et al. (2000)). The randomness of a $\mathscr{D}_{n,m}(\tau,c)$-digraph lies in the fact that the vertices are random with the joint distribution $F_{X,Y}$, but arcs (X_i, X_j) are deterministic functions of the random variable X_j and the random set $N(X_i,\tau,c)$ in the middle intervals and the random set $N_e(X_i,\tau)$ in the end intervals. In $\mathbb{R}$, the vertex-random PCD is a special case of *interval catch digraphs* (see, e.g., Sen et al. (1989) and Prisner (1994)). Furthermore, when $\tau = 1$ and $c = 1/2$ (i.e., $M_c = \left(Y_{(i-1)} + Y_{(i)}\right)/2$) we have $N(x,1,1/2) = B(x, r(x))$ for an x in a middle interval and $N_e(x,1) = B(x,r(x))$ for an x in an end interval where $r(x) = d(x,\mathcal{Y}_m) = \min_{y\in\mathcal{Y}_m} d(x,y)$ and the corresponding PCD is the CCCD of Priebe et al. (2001). See also Figure 1 (right).

3 Relative Density of Vertex-Random PCDs

Let $D_n = (\mathcal{V}, \mathcal{A})$ be a digraph with vertex set $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ and arc set $\mathcal{A}$ and let $|\cdot|$ stand for the set cardinality function. The relative density of the digraph D_n which is of order $|\mathcal{V}| = n \geq 2$, denoted $\rho(D_n)$, is defined as (Janson et al. (2000))

$$\rho(D_n) = \frac{|\mathcal{A}|}{n(n-1)}.$$

Thus $\rho(D_n)$ represents the ratio of the number of arcs in the digraph D_n to the number of arcs in the complete symmetric digraph of order n, which is $n(n-1)$. For $n \leq 1$, we set $\rho(D_n) = 0$, since there are no arcs. If D_n is a random digraph in which arcs result from a random process, then the *arc probability* between vertices v_i, v_j is $p_a(i,j) = P((v_i, v_j) \in \mathcal{A})$ for all $i \neq j$, $i,j = 1,2,\dots,n$.

Given $\mathcal{Y}_m = \{y_1, y_2, \ldots, y_m\}$, let $\mathcal{X}_n$ be a random sample from F_X and D_n be the PCD based on proximity region $N(\cdot)$ with vertices $\mathcal{X}_n$ and the arc set $\mathcal{A}$ is defined as $(X_i, X_j) \in \mathcal{A}$ if $X_j \in N(X_i)$. Let $h_{ij} := (g_{ij} + g_{ji})/2$ where $g_{ij} = I((X_i, X_j) \in \mathcal{A}) + \mathbf{I}(X_j \in N(X_i))$. Then we can rewrite the relative density as follows:

$$\rho(D_n) = \frac{2}{n(n-1)} \sum\sum_{i<j} h_{ij}.$$

Although the digraph is asymmetric, h_{ij} is defined as the average number of arcs between X_i and X_j in order to produce a symmetric kernel with finite variance (Lehmann (1988)). The relative density $\rho(D_n)$ is a random variable that depends on n, F, and $N(\cdot)$ (i.e., $\mathcal{Y}_m$). But $\mathbf{E}\left[\rho(D_n)\right] = \mathbf{E}\left[h_{12}\right] = p_a$ only depends on F and $N(\cdot)$. Furthermore,

$$0 \leq \mathbf{Var}\left[\rho(D_n)\right] = \frac{4}{n^2\,(n-1)^2} \mathbf{Var}\left[\sum\sum_{i<j} h_{ij}\right] = \frac{2}{n\,(n-1)} \mathbf{Var}\left[h_{12}\right] + \frac{4(n-2)}{n\,(n-1)} \mathbf{Cov}\left[h_{12}, h_{13}\right] \leq 1/4. \tag{6}$$

Hence $\rho(D_n)$, is a one-sample U-statistic of degree 2 and is an unbiased estimator of arc probability p_a. If, additionally, $\nu = \mathbf{Cov}\left[h_{ij}, h_{ik}\right] > 0$ for all $i \neq j \neq k$, $i, j, k \in \{1, 2, \ldots, n\}$, then a CLT for U-statistics (Lehmann (1988)) yields $\sqrt{n}\,\left[\rho(D_n) - p_a\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu)$ as $n \to \infty$, where $\xrightarrow{\mathcal{L}}$ stands for convergence in law and $\mathcal{N}(\mu, \sigma^2)$ stands for the normal distribution with mean μ and variance σ^2.

In Equation (6), we have

$$\begin{aligned}\mathbf{Var}[h_{ij}] &= \mathbf{Var}\left[h_{12}\right] = \mathbf{E}[(h_{ij})^2] - (\mathbf{E}[h_{ij}])^2 = \mathbf{E}[(g_{ij} + g_{ji})^2/4] - p_a^2 = \\ &(\mathbf{E}[g_{ij}] + 2\,\mathbf{E}[g_{ij}]\mathbf{E}[g_{ji}] + \mathbf{E}[g_{ji}])/4 - p_a^2 = (p_a + 2\,p_a + p_a)/4 - p_a^2 = (p_a - p_a^2)/2 = p_a\,(1 - p_a)\,/2\end{aligned}$$

and the covariance is

$$\mathbf{Cov}\left[h_{12}, h_{13}\right] = \mathbf{E}\left[h_{12}h_{13}\right] - \mathbf{E}\left[h_{12}\right]\mathbf{E}\left[h_{13}\right] = \mathbf{E}\left[h_{12}h_{13}\right] - p_a^2,$$

with

$$\begin{aligned}4\,\mathbf{E}\left[h_{12}h_{13}\right] &= \mathbf{E}[(g_{12} + g_{21})(g_{13} + g_{31})] = \mathbf{E}[g_{12}g_{13} + g_{12}g_{31} + g_{21}g_{13} + g_{21}g_{31}] \\ &= \mathbf{E}[\mathbf{I}(X_2 \in N(X_1)\mathbf{I}(X_3 \in N(X_1)) + \mathbf{I}(X_2 \in N(X_1)\mathbf{I}(X_1 \in N(X_3)) + \\ &\quad \mathbf{I}(X_1 \in N(X_2)\mathbf{I}(X_3 \in N(X_1))] + \mathbf{I}(X_1 \in N(X_2)\mathbf{I}(X_1 \in N(X_3))] \\ &= \mathbf{E}[\mathbf{I}(\{X_2, X_3\} \subset N(X_1)) + \mathbf{I}(X_2 \in N(X_1)\mathbf{I}(X_3 \in \Gamma_1(X_3, N)) + \\ &\quad \mathbf{I}(X_2 \in \Gamma_1(X_1)\mathbf{I}(X_3 \in N(X_1))] + \mathbf{I}(X_2 \in \Gamma_1(X_1, N)\mathbf{I}(X_3 \in \Gamma_1(X_1, N))] \\ &= P(\{X_2, X_3\} \subset N(X_1)) + 2\,P(X_2 \in N(X_1), X_3 \in \Gamma_1(X_1, N)) + P(\{X_2, X_3\} \subset \Gamma_1(X_1, N)).\end{aligned}$$

Then $\nu = \mathbf{Cov}(h_{ij}, h_{ik}) = \mathbf{E}[h_{ij}h_{ik}] - \mathbf{E}[h_{ij}]\mathbf{E}[h_{ik}] = \mathbf{E}[h_{ij}h_{ik}] - p_a^2 = \mathbf{E}[h_{12}h_{13}] - p_a^2 > 0$ iff

$$P(\{X_2, X_3\} \subset N(X_1)) + 2\,P(X_2 \in N(X_1), X_3 \in \Gamma_1(X_1, N)) + P(\{X_2, X_3\} \subset \Gamma_1(X_1, N)) > 4p_a^2.$$

Notice also that

$$\begin{aligned}\mathbf{E}[|h_{ij}|^3] &= \mathbf{E}[(g_{ij} + g_{ji})^3/8] = \mathbf{E}[g_{ij}^3 + 3\,g_{ij}^2 g_{ji} + 3\,g_{ij}g_{ji}^2 + g_{ji}^3]/8 = \mathbf{E}[g_{ij} + 3\,g_{ij}g_{ji} + 3\,g_{ij}g_{ji} + g_{ji}]/8 = \\ &(2\,\mathbf{E}[g_{ij}] + 6\,\mathbf{E}[g_{ij}]\mathbf{E}[g_{ji}])/8 = (p_a + 3\,p_a^2)/4 < \infty.\end{aligned}$$

Then for $\nu > 0$, the sharpest rate of convergence in the asymptotic normality of $\rho(D_n)$ is

$$\sup_{t \in \mathbb{R}} \left| P\left(\frac{\sqrt{n}(\rho(D_n) - p_a)}{\sqrt{4\,\nu}} \leq t\right) - \Phi(t) \right| \leq 8\,K\,p_a\,(4\,\nu)^{-3/2}\,n^{-1/2} = K\,\frac{p_a}{\sqrt{n\,\nu^3}} \tag{7}$$

where K is a constant and $\Phi(t)$ is the distribution function for the standard normal distribution (Callaert and Janssen (1978)).

In general a random digraph, just like a random graph, can be obtained by starting with a set of n vertices and adding arcs between them at random. We can consider the digraph counterpart of the Erdős–Rényi model for random graphs, denoted $D(n,p)$, in which every possible arc occurs independently with probability p (Erdős and Rényi (1959)). Notice that for the random digraph $D(n,p)$, the relative density of $D(n,p)$ is a U-statistic; however, the asymptotic distribution of its relative density is degenerate (with $\rho(D(n,p)) \xrightarrow{\mathcal{L}} p$, as $n \to \infty$) since the covariance term is zero due to the independence between the arcs.

Let $\mathcal{F}(\mathbb{R}) := \{F_{X,Y}$ on $\mathbb{R}$ with $P(X = Y) = 0$ and the marginals, F_X and F_Y, are non-atomic$\}$. In this article, we consider $\mathscr{D}_{n,m}(\tau,c)$-digraphs for which $\mathcal{X}_n$ and $\mathcal{Y}_m$ are random samples from F_X and F_Y, respectively, and the joint distribution of X,Y is $F_{X,Y} \in \mathcal{F}(\mathbb{R})$. Then the order statistics of $\mathcal{X}_n$ and $\mathcal{Y}_m$ are distinct with probability one. We call such digraphs as $\mathcal{F}(\mathbb{R})$*-random* $\mathscr{D}_{n,m}(\tau,c)$*-digraphs* and focus on the random variable $\rho(\mathscr{D}_{n,m}(\tau,c))$. For notational brevity, we use $\rho_{n,m}(\tau,c)$ instead of $\rho(\mathscr{D}_{n,m}(\tau,c))$. It is trivial to see that $0 \le \rho_{n,m}(\tau,c) \le 1$, and $\rho_{n,m}(\tau,c) > 0$ for nontrivial digraphs.

3.1 The Distribution of the Relative Density of $\mathcal{F}(\mathbb{R})$-random $\mathscr{D}_{n,m}(\tau,c)$-digraphs

Let $\mathcal{I}_i := \left(Y_{(i-1)}, Y_{(i)}\right)$, $\mathcal{X}_{[i]} := \mathcal{X}_n \cap \mathcal{I}_i$, and $\mathcal{Y}_{[i]} := \{Y_{(i-1)}, Y_{(i)}\}$ for $i = 1,2,\ldots,(m+1)$. Let $D_{[i]}(\tau,c)$ be the component of the random $\mathscr{D}_{n,m}(\tau,c)$-digraph induced by the pair $\mathcal{X}_{[i]}$ and $\mathcal{Y}_{[i]}$. Then we have a disconnected digraph with subdigraphs $D_{[i]}(\tau,c)$ for $i = 1,2,\ldots,(m+1)$ each of which might be null or itself disconnected. Let $\mathcal{A}_{[i]}$ be the arc set of $D_{[i]}(\tau,c)$, and $\rho_{[i]}(\tau,c)$ denote the relative density of $D_{[i]}(\tau,c)$; $n_i := \left|\mathcal{X}_{[i]}\right|$, and F_i be the density F_X restricted to $\mathcal{I}_i$ for $i \in \{1,2,\ldots,m+1\}$. Furthermore, let $M_c^{[i]} \in \mathcal{I}_i$ be the point so that it divides the interval $\mathcal{I}_i$ in ratios c and $1-c$ (i.e., length of the subinterval to the left of $M_c^{[i]}$ is $c \times 100$ % of the length of $\mathcal{I}_i$) for $i \in \{2,\ldots,m\}$. Notice that for $i \in \{2,\ldots,m\}$ (i.e., middle intervals), $D_{[i]}(\tau,c)$ is based on the proximity region $N(x,\tau,c)$ and for $i \in \{1, m+1\}$ (i.e., end intervals), $D_{[i]}(\tau,c)$ is based on the proximity region $N_e(x,\tau)$. Since we have at most $m+1$ subdigraphs that are disconnected, it follows that we have at most $n_{_T} := \sum_{i=1}^{m+1} n_i(n_i-1)$ arcs in the digraph $\mathscr{D}_{n,m}(\tau,c)$. Then we define the relative density for the entire digraph as

$$\rho_{n,m}(\tau,c) := \frac{|\mathcal{A}|}{n_{_T}} = \frac{\sum_{i=1}^{m+1} |\mathcal{A}_{[i]}|}{n_{_T}} = \frac{1}{n_{_T}} \sum_{i=1}^{m+1} (n_i(n_i-1))\rho_{_{[i]}}(\tau,c). \tag{8}$$

Since $\frac{n_i\,(n_i-1)}{n_{_T}} \ge 0$ for each i and $\displaystyle\sum_{i=1}^{m+1} \frac{n_i\,(n_i-1)}{n_{_T}} = 1$, it follows that $\rho_{n,m}(\tau,c)$ is a mixture of the $\rho_{_{[i]}}(\tau,c)$. We study the simpler random variable $\rho_{_{[i]}}(\tau,c)$ first. In the remaining of this section, the almost sure (a.s.) results follow from the fact that the marginal distributions F_X and F_Y are non-atomic.

Lemma 3.1. *Let $D_{[i]}(\tau,c)$ be the digraph induced by $\mathcal{X}$ points in the end intervals (i.e., $i \in \{1,(m+1)\}$) and $\rho_{_{[i]}}(\tau,c)$ be the corresponding relative density. For $\tau > 0$, if $n_i \le 1$, then $\rho_{_{[i]}}(\tau,c) = 0$. For $\tau \ge 1$, if $n_i > 1$, then $\rho_{_{[i]}}(\tau,c) \ge 1/2$ a.s.*

Proof: Let $i = m+1$ (i.e., consider the right end interval). For all $\tau > 0$, if $n_{m+1} \le 1$, then by definition $\rho_{_{[m+1]}}(\tau,c) = 0$. So, we assume $n_{m+1} > 1$. Let $\mathcal{X}_{[m+1]} = \{Z_1, Z_2, \ldots, Z_{n_{m+1}}\}$ and $Z_{(j)}$ be the corresponding order statistics. Then for $\tau \ge 1$, there is an arc from $Z_{(j)}$ to each $Z_{(k)}$ for $k < j$, with $j,k \in \{1,2,\ldots,n_{m+1}\}$ (and possibly to some other Z_l), since $N_e\left(Z_{(j)},\tau\right) = \left(Y_{(m)}, Z_{(j)} + \tau\left(Z_{(j)} - Y_{(m)}\right)\right)$ and so $Z_{(k)} \in N_e\left(Z_{(j)},\tau\right)$. So, there are at least $0 + 1 + 2 + \ldots + n_{m+1} - 1 = n_{m+1}(n_{m+1}-1)/2$ arcs in $D_{[m+1]}(\tau,c)$. Then $\rho_{_{[i]}}(\tau,c) \ge (n_{m+1}(n_{m+1}-1)/2)/(n_{m+1}(n_{m+1}-1)) = 1/2$. By symmetry, the same results hold for $i = 1$. ■

Using Lemma 3.1, we obtain the following lower bound for $\rho_{n,m}(\tau,c)$ for $\tau \ge 1$.

Theorem 3.2. *Let $D_{n,m}(\tau,c)$ be an $\mathcal{F}(\mathbb{R})$-random $\mathscr{D}_{n,m}(\tau,c)$-digraph with $n>0$, $m>0$ and k_1 and k_2 be two natural numbers defined as $k_1 := \sum_{i=2}^{m}(n_{i,1}(n_{i,1}-1)/2 + n_{i,2}(n_{i,2}-1)/2)$ and $k_2 := \sum_{i\in\{1,m+1\}} n_i(n_i-1)/2$, where $n_{i,1} := \left|\mathcal{X}_n \cap \left(Y_{(i-1)}, M_c^{[i]}\right)\right|$ and $n_{i,2} := \left|\mathcal{X}_n \cap \left(M_c^{[i]}, Y_{(i)}\right)\right|$. Then for $\tau \geq 1$, we have $(k_1+k_2)/n_T \leq \rho_{n,m}(\tau,c) \leq 1$ a.s.*

Proof: For $i \in \{1,(m+1)\}$, we have k_2 as in Lemma 3.1. Let $i \in \{2,3,\ldots,m\}$ and $\mathcal{X}_{i,1} := \mathcal{X}_{[i]} \cap \left(Y_{(i-1)}, M_c^{[i]}\right) = \{U_1,U_2,\ldots,U_{n_{i,1}}\}$, and $\mathcal{X}_{i,2} := \mathcal{X}_{[i]} \cap \left(M_c^{[i]}, Y_{(i)}\right) = \{V_1,V_2,\ldots,V_{n_{i,2}}\}$. Furthermore, let $U_{(j)}$ and $V_{(k)}$ be the corresponding order statistics. For $\tau \geq 1$, there is an arc from $U_{(j)}$ to $U_{(k)}$ for $k<j$, $j,k \in \{1,2,\ldots,n_{i,1}\}$ and possibly to some other U_l, and similarly there is an arc from $V_{(j)}$ to $V_{(k)}$ for $k>j$, $j,k \in \{1,2,\ldots,n_{i,2}\}$ and possibly to some other V_l. Thus there are at least $\frac{n_{i,1}(n_{i,1}-1)}{2} + \frac{n_{i,2}(n_{i,2}-1)}{2}$ arcs in $D_{[i]}(\tau,c)$. Hence $\rho_{n,m}(\tau,c) \geq (k_1+k_2)/n_T$. ■

Theorem 3.3. *For $i=1,2,3,\ldots,m+1$, $\tau=\infty$, and $n_i>0$, we have $\rho_{[i]}(\tau=\infty,c) = \mathbf{I}(n_i>1)$ and $\rho_{n,m}(\tau=\infty,c)=1$ a.s.*

Proof: For $\tau=\infty$, if $n_i \leq 1$, then $\rho_{[i]}(\tau=\infty,c)=0$. So we assume $n_i>1$ and let $i=m+1$. Then $N_e(x,\infty) = \left(Y_{(m)},\infty\right)$ for all $x \in \left(Y_{(m)},\infty\right)$. Hence $D_{[m+1]}(\infty,c)$ is a complete symmetric digraph of order n_{m+1}, which implies $\rho_{[m+1]}(\tau=\infty,c)=1$. By symmetry, the same holds for $i=1$. For $i \in \{2,3,\ldots,m\}$ and $n_i>1$, we have $N(x,\infty,c)=\mathcal{I}_i$ for all $x \in \mathcal{I}_i$, hence $D_{[i]}(\infty,c)$ is a complete symmetric digraph of order n_i, which implies $\rho_{[i]}(\infty,c)=1$. Then $\rho_{n,m}(\infty,c) = \sum \frac{n_i(n_i-1)\rho_{[i]}(\infty,c)}{n_T} = 1$, since when $n_i \leq 1$, n_i has no contribution to n_T, and when $n_i>1$, we have $\rho_{[i]}(\infty,c)=1$. ■

4 The Distribution of the Relative Density of Central Similarity PCDs for Uniform Data

Let $-\infty<\delta_1<\delta_2<\infty$, $\mathcal{Y}_m$ be a random sample from non-atomic F_Y with support $\mathcal{S}(F_Y) \subseteq (\delta_1,\delta_2)$, and $\mathcal{X}_n = \{X_1,X_2,\ldots,X_n\}$ be a random sample from $F_X = \mathcal{U}(\delta_1,\delta_2)$, the uniform distribution on (δ_1,δ_2). So we have $F_{X,Y} \in \mathcal{F}(\mathbb{R})$. Assuming we have the realization of $\mathcal{Y}_m$ as $\mathcal{Y}_m = \{y_1,y_2,\ldots,y_m\} = \{y_{(1)},y_{(2)},\ldots,y_{(m)}\}$ with $\delta_1 < y_{(1)} < y_{(2)} < \ldots < y_{(m)} < \delta_2$, we let $y_{(0)} := \delta_1$ and $y_{(m+1)} := \delta_2$. Then it follows that the distribution of X_i restricted to $\mathcal{I}_i$ is $F_X|_{\mathcal{I}_i} = \mathcal{U}(\mathcal{I}_i)$. We call such digraphs as *$\mathcal{U}(\delta_1,\delta_2)$-random $\mathscr{D}_{n,m}(\tau,c)$-digraphs* and provide the distribution of their relative density for the whole range of τ and c. We first present a "scale invariance" result for central similarity PCDs. This invariance property will simplify the notation in our subsequent analysis by allowing us to consider the special case of the unit interval $(0,1)$.

Theorem 4.1. *(Scale Invariance Property) Suppose $\mathcal{X}_n$ is a set of iid random variables from $\mathcal{U}(\delta_1,\delta_2)$ where $\delta_1<\delta_2$ and $\mathcal{Y}_m$ is set of m distinct $\mathcal{Y}$ points in (δ_1,δ_2). Then for any $\tau>0$, the distribution of $\rho_{[i]}(\tau,c)$ is independent of $\mathcal{Y}_{[i]}$ (and hence of the restricted support interval $\mathcal{I}_i$) for all $i \in \{1,2,\ldots,m+1\}$.*

Proof: Let $\delta_1<\delta_2$ and $\mathcal{Y}_m$ be as in the hypothesis. Any $\mathcal{U}(\delta_1,\delta_2)$ random variable can be transformed into a $\mathcal{U}(0,1)$ random variable by $\phi(x) = (x-\delta_1)/(\delta_2-\delta_1)$, which maps intervals $(t_1,t_2) \subseteq (\delta_1,\delta_2)$ to intervals $\left(\phi(t_1),\phi(t_2)\right) \subseteq (0,1)$. That is, if $X \sim \mathcal{U}(\delta_1,\delta_2)$, then we have $\phi(X) \sim \mathcal{U}(0,1)$ and $P(X \in (t_1,t_2)) = P(\phi(X) \in \left(\phi(t_1),\phi(t_2)\right)$ for all $(t_1,t_2) \subseteq (\delta_1,\delta_2)$. The distribution of $\rho_{[i]}(\tau,c)$ is obtained by calculating such probabilities. So, without loss of generality, we can assume $\mathcal{X}_{[i]}$ is a set of iid random variables from the $\mathcal{U}(0,1)$ distribution. That is, the distribution of $\rho_{[i]}(\tau,c)$ does not depend on $\mathcal{Y}_{[i]}$ and hence does not depend on the restricted support interval $\mathcal{I}_i$. ■

Note that scale invariance of $\rho_{[i]}(\tau=\infty,c)$ follows trivially for all $\mathcal{X}_n$ from any F_X with support in (δ_1,δ_2) with $\delta_1<\delta_2$, since for $\tau=\infty$, we have $\rho_{[i]}(\tau=\infty,c)=1$ a.s. for non-atomic F_X.

Based on Theorem 4.1, we may assume each $\mathcal{I}_i$ as the unit interval $(0,1)$ for uniform data. Then the central similarity proximity region for $x \in (0,1)$ with parameters $c \in (0,1)$ and $\tau > 0$ have the following forms. If $x \in \mathcal{I}_i$ for $i \in \{2,\dots,m\}$ (i.e., in the middle intervals), when transformed under $\phi(\cdot)$ to $(0,1)$, we have

$$N(x,\tau,c) = \begin{cases} (x\,(1-\tau), x\,(c+(1-c)\,\tau)/c) \cap (0,1) & \text{if } x \in (0,c), \\ (x - c\,\tau\,(1-x)/(1-c), x + (1-x)\,\tau) \cap (0,1) & \text{if } x \in (c,1). \end{cases} \tag{9}$$

In particular, for $\tau \in (0,1)$, we have

$$N(x,\tau,c) = \begin{cases} (x\,(1-\tau), x\,(c+(1-c)\,\tau)/c) & \text{if } x \in (0,c), \\ (x - c\,\tau\,(1-x)/(1-c), x + (1-x)\,\tau) & \text{if } x \in (c,1) \end{cases} \tag{10}$$

and for $\tau \geq 1$, we have

$$N(x,\tau,c) = \begin{cases} (0, x\,(c+(1-c)\,\tau)/c) & \text{if } x \in \left(0, \frac{c}{c+(1-c)\,\tau}\right), \\ (0,1) & \text{if } x \in \left(\frac{c}{c+(1-c)\,\tau}, \frac{c\,\tau}{1-c+c\,\tau}\right), \\ (x - c\,\tau\,(1-x)/(1-c), 1) & \text{if } x \in \left(\frac{c\,\tau}{1-c+c\,\tau}, 1\right). \end{cases} \tag{11}$$

and $N(x=c,\tau,c)$ is arbitrarily taken to be one of $(x\,(1-\tau), x\,(c+(1-c)\,\tau)/c) \cap (0,1)$ or $(x - c\,\tau\,(1-x)/(1-c), x + (1-x)\,\tau) \cap (0,1)$. This special case of "$X = c$" happens with probability zero for uniform X.

If $x \in \mathcal{I}_1$ (i.e., in the left end interval), when transformed under $\phi(\cdot)$ to $(0,1)$, we have $N_e(x,\tau) = (\max(0, x - \tau\,(1-x)), \min(1, x + \tau\,(1-x))$; and if $x \in \mathcal{I}_{m+1}$ (i.e., in the right end interval), when transformed under $\phi(\cdot)$ to $(0,1)$, we have $N_e(x,\tau) = (\max(0, x\,(1-\tau)), \min(1, x\,(1+\tau)))$.

Notice that each subdigraph $D_{[i]}(\tau,c)$ is itself a $\mathcal{U}(\mathcal{I}_i)$-random $\mathscr{D}_{n,2}(\tau,c)$-digraph. The distribution of the relative density of $D_{[i]}(\tau,c)$ is given in the following result.

Theorem 4.2. *Let $\rho_{[i]}(\tau,c)$ be the relative density of subdigraph $D_{[i]}(\tau,c)$ of the central similarity PCD based on uniform data in (δ_1,δ_2) where $\delta_1 < \delta_2$ and $\mathcal{Y}_m$ be a set of m distinct $\mathcal{Y}$ points in (δ_1,δ_2). Then for $\tau \in (0,\infty)$, as $n_i \to \infty$, we have*

(i) for $i \in \{2,\dots,m\}$, $\sqrt{n_i}\,\left[\rho_{[i]}(\tau,c) - \mu(\tau,c)\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu(\tau,c))$, where $\mu(\tau,c) = \mathbf{E}\left[\rho_{[i]}(\tau,c)\right]$ is the arc probability and $\nu(\tau,c) = \mathbf{Cov}[h_{12},h_{12}]$ in the middle intervals, and

(ii) for $i \in \{1,m+1\}$, $\sqrt{n_i}\,\left[\rho_{[i]}(\tau,c) - \mu_e(\tau)\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu_e(\tau))$, where $\mu_e(\tau) = \mathbf{E}\left[\rho_{[i]}(\tau,c)\right]$ is the arc probability and $\nu_e(\tau) = \mathbf{Cov}[h_{12},h_{12}]$ in the end intervals.

Proof: (i) Let $i \in \{2,\dots,m\}$ (i.e., $\mathcal{I}_i$ be a middle interval). By the scale invariance for uniform data (see Theorem 4.1), a middle interval can be assumed to be the unit interval $(0,1)$. The mean of the asymptotic distribution of $\rho_{[i]}(\tau,c)$ is computed as follows.

$$\mathbf{E}[\rho_{[i]}(\tau,c)] = \mathbf{E}[h_{12}] = P(X_2 \in N(X_1,\tau,c)) = \mu(\tau,c)$$

which is the arc probability. And the asymptotic variance of $\rho_{[i]}(\tau,c)$ is $\mathbf{Cov}[h_{12},h_{13}] = 4\,\nu(\tau,c)$. For $\tau \in (0,\infty)$, since $2h_{12} = \mathbf{I}(X_2 \in N(X_1,\tau,c)) + \mathbf{I}(X_1 \in N(X_2,\tau,c))$ is the number of arcs between X_1 and X_2 in the PCD, h_{12} tends to be high if the proximity region $N(X_1,\tau,c)$ is large. In such a case, h_{13} tends to be high also. That is, h_{12} and h_{13} tend to be high and low together. So, for $\tau \in (0,\infty)$, we have $\nu(\tau,c) > 0$. Hence asymptotic normality follows.

(ii) In an end interval, the mean of the asymptotic distribution of $\rho_{[i]}(\tau,c)$ is

$$\mathbf{E}[\rho_{[i]}(\tau,c)] = \mathbf{E}[h_{12}] = P(X_2 \in N_e(X_1,\tau)) = \mu_e(\tau)$$

the asymptotic variance of $\rho_{[i]}(\tau, c)$ is $\mathbf{Cov}[h_{12}, h_{13}] = 4\,\nu_e(\tau)$. For $\tau \in (0, \infty)$, as in (i), we have $\nu_e(\tau) > 0$. Hence asymptotic normality follows. ■

Let $P_{2N} := P(\{X_2, X_3\} \subset N(X_1, \tau, c))$, $P_{NG} := P(X_2 \in N(X_1, \tau, c), X_3 \in \Gamma_1(X_1, \tau, c))$, and $P_{2G} := P(\{X_2, X_3\} \subset \Gamma_1(X_1, \tau, c))$. Then

$$\mathbf{Cov}[h_{12}, h_{13}] = \mathbf{E}[h_{12}h_{13}] - \mathbf{E}[h_{12}]\mathbf{E}[h_{13}] = \mathbf{E}[h_{12}h_{13}] - \mu(\tau, c)^2 = (P_{2N} + 2\,P_{NG} + P_{2G})/4 - \mu(\tau, c)^2,$$

since

$$4\,\mathbf{E}[h_{12}h_{13}] = P(\{X_2, X_3\} \subset N(X_1, \tau, c)) + 2\,P(X_2 \in N(X_1, \tau, c), X_3 \in \Gamma_1(X_1, \tau, c)) + P(\{X_2, X_3\} \subset \Gamma_1(X_1, \tau, c)) = P_{2N} + 2\,P_{NG} + P_{2G}.$$

Similarly, let $P_{2N,e} := P(\{X_2, X_3\} \subset N_e(X_1, \tau))$, $P_{NG,e} := P(X_2 \in N_e(X_1, \tau), X_3 \in \Gamma_{1,e}(X_1, \tau))$, and $P_{2G,e} := P(\{X_2, X_3\} \subset \Gamma_{1,e}(X_1, \tau))$. Then

$$\mathbf{Cov}[h_{12}, h_{13}] = (P_{2N,e} + 2\,P_{NG,e} + P_{2G,e})/4 - \mu_e(\tau)^2.$$

For $\tau = \infty$, we have $N(x, \infty, c) = \mathcal{I}_i$ for all $x \in \mathcal{I}_i$ with $i \in \{2, \ldots, m\}$ and $N_e(x, \infty) = \mathcal{I}_i$ for all $x \in \mathcal{I}_i$ with $i \in \{1, m+1\}$. Then for $i \in \{2, \ldots, m\}$

$$\mathbf{E}\left[\rho_{[i]}(\infty, c)\right] = \mathbf{E}\left[h_{12}\right] = \mu(\infty, c) = P(X_2 \in N(X_1, \infty, c) = P(X_2 \in \mathcal{I}_i) = 1.$$

On the other hand, $4\,\mathbf{E}\left[h_{12}h_{13}\right] = P(\{X_2, X_3\} \subset N(X_1, \infty, c)) + 2\,P(X_2 \in N(X_1, \infty, c), X_3 \in \Gamma_1(X_1, \infty, c)) + P(\{X_2, X_3\} \subset \Gamma_1(X_1, \infty, c)) = (1 + 2 + 1)$. Hence $\mathbf{E}\left[h_{12}h_{13}\right] = 1$ and so $\nu(\infty, c) = 0$. Similarly, for $i \in \{1, m+1\}$, we have $\mu_e(\infty) = 1$ and $\nu_e(\infty) = 0$. Therefore, the CLT result does not hold for $\tau = \infty$. Furthermore, $\rho_{[i]}(\tau = \infty, c) = 1$ a.s.

By Theorem 4.2, we have $\nu(\tau, c) > 0$ (and $\nu_e(\tau) > 0$) iff $P_{2N} + 2\,P_{NG} + P_{2G} > 4\,\mu(\tau, c)^2$ (and $P_{2N,e} + 2\,P_{NG,e} + P_{2G,e} > 4\,\mu_e(\tau)^2$).

Remark 4.3. **The Joint Distribution of** (h_{12}, h_{13}): The pair (h_{12}, h_{13}) is a bivariate discrete random variable with nine possible values such that

$$(2\,h_{12}, 2\,h_{13}) \in \{(0,0), (0,1), (0,2), (1,0), (1,1), (1,2), (2,0), (2,1), (2,2)\}.$$

Then finding the joint distribution of (h_{12}, h_{13}) is equivalent to finding the joint probability mass function of (h_{12}, h_{13}). Hence the joint distribution of (h_{12}, h_{13}) can be found by calculating the probabilities such as $P((h_{12}, h_{13}) = (0,0)) = P(\{X_2, X_3\} \subset \mathcal{I}_i \setminus (N(X_1, \tau, c) \cup \Gamma_1(X_1, \tau, c)))$. □

4.1 The Distribution of Relative Density of $\mathcal{U}(y_1, y_2)$-random $\mathscr{D}_{n,2}(\tau, c)$-digraphs

In the special case of $m = 2$ with $\mathcal{Y}_2 = \{y_1, y_2\}$ and $\delta_1 = y_1 < y_2 = \delta_2$, we have only one middle interval and the two end intervals are empty. In this section, we consider the relative density of central similarity PCD based on uniform data in (y_1, y_2). By Theorems 4.1 and 4.2, the asymptotic distribution of any $\rho_{[i]}(\tau, c)$ for the middle intervals for $m > 2$ will be identical to the asymptotic distribution of $\mathcal{U}(y_1, y_2)$-random $\mathscr{D}_{n,2}(\tau, c)$-digraph.

First we consider the simplest case of $\tau = 1$ and $c = 1/2$. By Theorem 4.1, without loss of generality, we can assume (y_1, y_2) to be the unit interval $(0, 1)$. Then $N(x, 1, 1/2) = B(x, r(x))$ where $r(x) = \min(x, 1 - x)$ for $x \in (0, 1)$. Hence central similarity PCD based on $N(x, 1, 1/2)$ is equivalent to the CCCD of Priebe et al. (2001). Moreover, we have $\Gamma_1(X_1, 2, 1/2) = (X_1/2, (1 + X_1)/2)$.

Theorem 4.4. *As $n \to \infty$, we have $\sqrt{n}\,[\rho_n(1,1/2) - \mu(1,1/2)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu(1,1/2))$, where $\mu(1,1/2) = 1/2$ and $4\,\nu(1,1/2) = 1/12$.*

Proof: By symmetry, we only consider $X_1 \in (0,1/2)$. Notice that for $x \in (0,1/2)$, we have $N(x,1,1/2) = (0,2\,x)$ and $\Gamma_1(x,1,1/2) = (x/2,(1+x)/2)$. Hence $\mu(1,1/2) = P(X_2 \in N(X_1,1,1/2)) = 2\,P(X_2 \in N(X_1,1,1/2), X_1 \in (0,1/2))$ by symmetry. Here

$$P(X_2 \in N(X_1,1,1/2), X_1 \in (0,1/2)) = P(X_2 \in (0,2x_1), X_1 \in (0,1/2))$$
$$= \int_0^{1/2}\int_0^{2x_1} f_{1,2}(x_1,x_2)dx_2dx_1 = \int_0^{1/2}\int_0^{2x_1} 1dx_2dx_1 = \int_0^{1/2} 2x_1dx_1 = x_1^2\big|_0^{1/2} = 1/4.$$

Then $\mu(1,1/2) = 2\,(1/4) = 1/2$.

For $\mathbf{Cov}(h_{12},h_{13})$, we need to calculate P_{2N}, P_{NG}, and P_{2G}. The probability

$$P_{2N} = P(\{X_2,X_3\} \subset N(X_1,1,1/2)) = 2\,P(\{X_2,X_3\} \subset N(X_1,1,1/2), X_1 \in (0,1/2))$$

and $P(\{X_2,X_3\} \subset N(X_1,1,1/2), X_1 \in (0,1/2)) = \int_0^{1/2}(2x_1)^2dx_1 = 1/6$. So $P_{2N} = 2\,(1/6) = 1/3$.

$P_{NG} = 2\,P(X_2 \in N(X_1,1,1/2), X_3 \in \Gamma_1(X_1,1,1/2), X_1 \in (0,1/2))$ and

$$P(X_2 \in N(X_1,1,1/2), X_3 \in \Gamma_1(X_1,1,1/2), X_1 \in (0,1/2)) = \int_0^{1/2}(2x_1)(1/2)dx_1 = 1/8.$$

Then $P_{NG} = 2\,(1/8) = 1/4$.

Finally, we have $P_{2G} = 2\,P(\{X_2,X_3\} \subset \Gamma_1(X_1,1,1/2), X_1 \in (0,1/2))$ and $P(\{X_2,X_3\} \subset \Gamma_1(X_1,1,1/2), X_1 \in (0,1/2)) = \int_0^{1/2}(1/4)dx_1 = 1/8$. So $P_{2G} = 2\,(1/8) = 1/4$.

Therefore $4\,\mathbf{E}[h_{12}h_{13}] = 1/3 + 2\,(1/4) + 1/4 = 13/12$. Hence $4\,\nu(1,1/2) = 4\,\mathbf{Cov}[h_{12},h_{13}] = 13/12 - 4(1/2)^2 = 1/12$. ■

The sharpest rate of convergence in Theorem 4.4 is $K\,\frac{\mu(2,1/2)}{\sqrt{n\,\nu(2,1/2)^3}} = 12\sqrt{3}\,\frac{K}{\sqrt{n}}$.

Next we consider the more general case of $\tau = 1$ and $c \in (0,1)$. For $x \in (0,1)$, the proximity region has the following form:

$$N(x,1,c) = \begin{cases} (0,x/c) & \text{if } x \in (0,c), \\ ((x-c)/(1-c),1) & \text{if } x \in (c,1), \end{cases} \tag{12}$$

and the Γ_1-region is $\Gamma_1(x,1,c) = (c\,x,(1-c)\,x + c)$.

Theorem 4.5. *As $n \to \infty$, for $c \in (0,1)$, we have $\sqrt{n}\,[\rho_{n,2}(1,c) - \mu(1,c)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu(1,c))$, where $\mu(1,c) = 1/2$ and $4\,\nu(1,c) = c\,(1-c)/3$.*

Proof is provided in Appendix 1. See Figure 2 for $4\,\nu(1,c)$ with $c \in (0,1/2)$. Notice that $\mu(1,c)$ is constant (i.e., independent of c) and $\nu(1,c)$ is symmetric around $c = 1/2$ with $\nu(1,c) = \nu(1,1-c)$. Notice also that for $c = 1/2$, we have $\mu(1,c = 1/2) = 1/2$, and $4\,\nu(1,c = 1/2) = 1/12$, hence as $c \to 1/2$, the distribution of $\rho_{n,2}(1,c)$ converges to the one in Theorem 4.4. Furthermore, the sharpest rate of convergence in Theorem 4.5 is

$$K\,\frac{\mu(1,c)}{\sqrt{n\,\nu(1,c)^3}} = \frac{3\sqrt{3}}{2\,\sqrt{c^3\,(1-c)^3}}\,\frac{K}{\sqrt{n}} \tag{13}$$

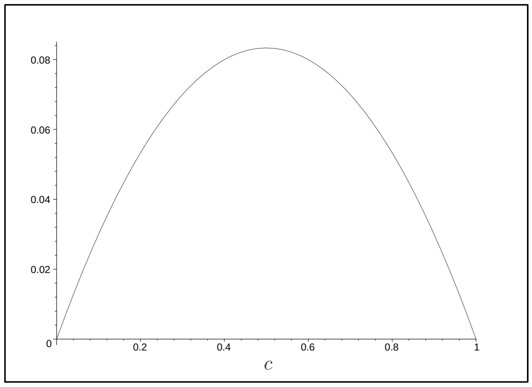

Figure 2: The plot of the asymptotic variance $4\,\nu(1,c)$ as a function of c for $c \in (0,1)$.

and is minimized at $c = 1/2$ (which can easily be verified).

Next we consider the case of $\tau > 0$ and $c = 1/2$. By symmetry, we only consider $X_1 \in (0,1/2)$. For $x \in (0,1/2)$, the proximity region for $\tau \in (0,1)$ is

$$N(x,\tau,1/2) = \begin{cases} (x\,(1-\tau), x\,(1+\tau)) & \text{if } x \in (0,1/2), \\ (x-(1-x)\,\tau, x+(1-x)\,\tau) & \text{if } x \in (1/2,1), \end{cases} \tag{14}$$

and for $\tau \geq 1$

$$N(x,\tau,1/2) = \begin{cases} (0, x\,(1+\tau)) & \text{if } x \in (0,1/(1+\tau)), \\ (0,1) & \text{if } x \in (1/(1+\tau),\tau/(1+\tau)), \\ (x-(1-x)\,\tau, 1) & \text{if } x \in (\tau/(1+\tau),1). \end{cases} \tag{15}$$

And the Γ_1-region for $\tau \in (0,1)$ is

$$\Gamma_1(x,\tau,1/2) = \begin{cases} (x/(1+\tau), x/(1-\tau)) & \text{if } x \in (0,(1-\tau)/2), \\ (x/(1+\tau), (x+\tau)/(1+\tau)) & \text{if } x \in ((1-\tau)/2,(1+\tau)/2), \\ ((x-\tau)/(1-\tau), (x+\tau)/(1+\tau)) & \text{if } x \in ((1+\tau)/2,1), \end{cases} \tag{16}$$

and for $\tau \geq 1$, we have $\Gamma_1(x,\tau,1/2) = (x/(1+\tau), (x+\tau)/(1+\tau))$.

Theorem 4.6. *For $\tau \in (0,\infty)$, we have $\sqrt{n}\,[\rho_{n,2}(\tau,1/2) - \mu(\tau,1/2)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu(\tau,1/2))$ as $n \to \infty$, where*

$$\mu(\tau,1/2) = \begin{cases} \tau/2 & \text{if } 0 < \tau < 1, \\ \tau/(\tau+1) & \text{if } \tau \geq 1, \end{cases} \tag{17}$$

and

$$4\,\nu(\tau,1/2) = \begin{cases} \frac{\tau^2\,(1+2\,\tau-\tau^2-\tau^3)}{3\,(\tau+1)^2} & \text{if } 0 < \tau < 1, \\ \frac{2\,\tau-1}{3\,(\tau+1)^2} & \text{if } \tau \geq 1. \end{cases} \tag{18}$$

Proof is provided in Appendix 1. See Figure 3 for the plots of $\mu(\tau,1/2)$ and $4\,\nu(\tau,1/2)$. Notice that $\lim_{\tau\to\infty}\nu(\tau,1/2) = 0$, so the CLT result fails for $\tau = \infty$. Furthermore, $\lim_{\tau\to 0}\nu(\tau,1/2) = 0$. For $\tau = 1$, we have $\mu(\tau = 1, c = 1/2) = 1/2$, and $4\,\nu(\tau = 1, c = 1/2) = 1/12$; hence as $\tau \to 1$, the distribution of $\rho_{n,2}(\tau,1/2)$ converges to the one in Theorem 4.4. Furthermore, the sharpest rate of convergence in Theorem 4.6 is

$$K\,\frac{\mu(\tau,1/2)}{\sqrt{n\,\nu(\tau,1/2)^3}} = \frac{K}{\sqrt{n}} \begin{cases} \frac{27\,\tau}{2}\left(\frac{\left(6\,\tau+3-3\,\tau^3-3\,\tau^2\right)\tau^2}{(\tau+1)^2}\right)^{-3/2} & \text{if } 0 < \tau < 1, \\ \frac{3\,\sqrt{3}\,\tau}{\tau+1}\left(\frac{2\,\tau-1}{(\tau+1)^2}\right)^{-3/2} & \text{if } \tau \geq 1. \end{cases} \tag{19}$$

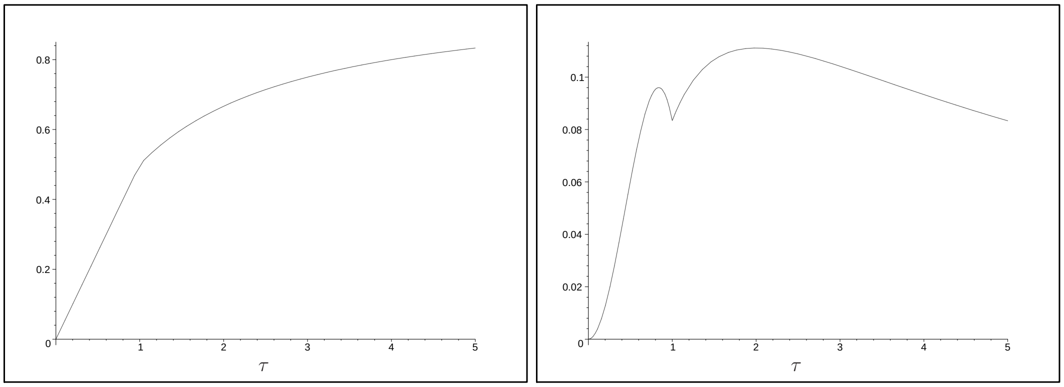

Figure 3: The plots of the asymptotic mean $\mu(\tau, 1/2)$ (left) and the variance $4\,\nu(\tau, 1/2)$ (right) as a function of τ for $\tau \in (0, 5]$.

and is minimized at $\tau \approx .73$ which is found by setting the first derivative of this rate with respect to τ to zero and solving for τ numerically. We also checked the plot of $\mu(\tau, 1/2)/\sqrt{\nu(\tau, 1/2)^3}$ (not presented) and verified that this is where the global minimum is attained.

Finally, we consider the most general case of $\tau > 0$ and $c \in (0, 1/2)$. For $\tau \in (0, 1)$, the proximity region is

$$N(x, \tau, c) = \begin{cases} \left(x\,(1-\tau), x\,\left(1 + \frac{(1-c)\,\tau}{c}\right)\right) & \text{if } x \in (0, c), \\ \left(x - \frac{c\,\tau\,(1-x)}{1-c}, x + (1-x)\,\tau\right) & \text{if } x \in (c, 1), \end{cases} \tag{20}$$

and the Γ_1-region is

$$\Gamma_1(x, \tau, c) = \begin{cases} \left(\frac{c\,x}{c+(1-c)\,\tau}, \frac{x}{1-\tau}\right) & \text{if } x \in (0, c\,(1-\tau)), \\ \left(\frac{c\,x}{c+(1-c)\,\tau}, \frac{x\,(1-c)+c\,\tau}{1-c+c\,\tau}\right) & \text{if } x \in (c\,(1-\tau), c\,(1-\tau)+\tau), \\ \left(\frac{x-\tau}{1-\tau}, \frac{x\,(1-c)+c\,\tau}{1-c+c\,\tau}\right) & \text{if } x \in (c\,(1-\tau)+\tau, 1). \end{cases} \tag{21}$$

For $\tau \geq 1$, the proximity region is

$$N(x, \tau, c) = \begin{cases} \left(0, x\,\left(1 + \frac{(1-c)\,\tau}{c}\right)\right) & \text{if } x \in \left(0, \frac{c}{c+(1-c)\,\tau}\right), \\ (0, 1) & \text{if } x \in \left(\frac{c}{c+(1-c)\,\tau}, \frac{c\,\tau}{1-c+c\,\tau}\right), \\ \left(x - \frac{c\,\tau\,(1-x)}{1-c}, 1\right) & \text{if } x \in \left(\frac{c\,\tau}{1-c+c\,\tau}, 1\right), \end{cases} \tag{22}$$

and the Γ_1-region is

$$\Gamma_1(x, \tau, c) = \left(\frac{c\,x}{c+(1-c)\,\tau}, \frac{x\,(1-c)+c\,\tau}{1-c+c\,\tau}\right). \tag{23}$$

Theorem 4.7. *For $\tau \in (0, \infty)$, we have $\sqrt{n}\,[\rho_{n,2}(\tau, c) - \mu(\tau, c)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu(\tau, c))$, as $n \to \infty$, where $\mu(\tau, c) = \mu_1(\tau, c)\,\mathbf{I}(0 < c \leq 1/2) + \mu_2(\tau, c)\,\mathbf{I}(1/2 \leq c < 1)$ and $\nu(\tau, c) = \nu_1(\tau, c)\,\mathbf{I}(0 < c \leq 1/2) + \nu_2(\tau, c)\,\mathbf{I}(1/2 \leq c < 1)$. For $0 < c \leq 1/2$,*

$$\mu_1(\tau, c) = \begin{cases} \frac{\tau}{2} & \text{if } 0 < \tau < 1, \\ \frac{\tau\,(1+2\,c\,(\tau-1)(1-c))}{2\,(c\,\tau-c+1)(\tau+c-c\,\tau)} & \text{if } \tau \geq 1, \end{cases} \tag{24}$$

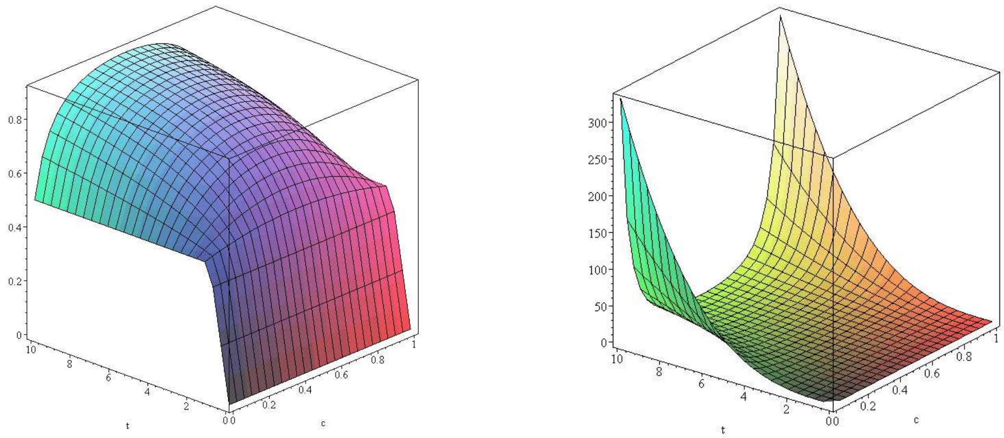

Figure 4: The surface plots of the asymptotic mean $\mu(t,c)$ (left) and the variance $4\,\nu(t,c)$ (right) as a function of t and c for $t \in (0,10]$ and $c \in (0,1)$, respectively.

and

$$4\,\nu_1(\tau,c) = \begin{cases} \kappa_1(\tau,c) & \textit{if } 0 < \tau < 1, \\ \kappa_2(\tau,c) & \textit{if } \tau \geq 1, \end{cases} \tag{25}$$

where

$$\kappa_1(\tau,c) = \frac{\tau^2\left(c^2\,\tau^3 - 3\,c^2\,\tau^2 - c\,\tau^3 + 2\,c^2\,\tau + 3\,c\,\tau^2 - c^2 - 2\,c\,\tau - \tau^2 + c + \tau\right)}{3\,\left(c\,\tau - c + 1\right)\left(c + \tau - c\,\tau\right)},$$

and

$$\kappa_2(\tau,c) = \Big[c(1-c)\big(2\,c^4\,\tau^5 - 7\,c^4\,\tau^4 - 4\,c^3\,\tau^5 + 8\,c^4\,\tau^3 + 14\,c^3\,\tau^4 + 3\,c^2\,\tau^5 - 2\,c^4\,\tau^2 - 16\,c^3\,\tau^3 - 7\,c^2\,\tau^4 - c\,\tau^5 - 2\,c^4\,\tau + 4\,c^3\,\tau^2 + 12\,c^2\,\tau^3 + c^4 + 4\,c^3\,\tau - 6\,c^2\,\tau^2 - 4\,c\,\tau^3 - 2\,c^3 - 3\,c^2\,\tau + 4\,c\,\tau^2 + c^2 + c\,\tau - \tau^2\big)\Big] \Big/ \Big[3\,\left(c\,\tau - c + 1\right)^3\left(c\,\tau - c - \tau\right)^3\Big].$$

And for $1/2 \leq c < 1$, we have $\mu_2(\tau,c) = \mu_1(\tau,1-c)$ and $\nu_2(\tau,c) = \nu_1(\tau,1-c)$.

Proof is provided in Appendix 1. See Figure 4 for the plots of $\mu(\tau,c)$ and $4\,\nu(\tau,c)$. Notice that $\lim_{\tau\to\infty}\nu(\tau,c) = 0$, so the CLT result fails for $\tau = \infty$. Furthermore, $\lim_{\tau\to 0}\nu(\tau,c) = 0$. For $\tau = 1$ and $c = 1/2$, we have $\mu(\tau = 1, c = 1/2) = 1/2$, and $4\,\nu(\tau = 1, c = 1/2) = 1/12$, hence as $\tau \to 1$ and $c \to 1/2$, the distribution of $\rho_{n,2}(\tau,c)$ converges to the one in Theorem 4.4. The sharpest rate of convergence in Theorem 4.7 is $K\,\frac{\mu(\tau,c)}{\sqrt{n\,\nu(\tau,c)^3}}$ (the explicit form not presented) and is minimized at $\tau \approx 1.55$ and $c \approx 0.5$ which is found by setting the first order partial derivatives of this rate with respect to τ and c to zero and solving for τ and c numerically. We also checked the surface plot of this rate (not presented) and verified that this is where the global minimum is attained.

4.2 The Case of End Intervals: Relative Density for $\mathcal{U}\left(\delta_1, y_{(1)}\right)$ or $\mathcal{U}\left(y_{(m)}, \delta_2\right)$ Data

Recall that with $m \geq 1$ for the end intervals, $\mathcal{I}_1 = \left(\delta_1, y_{(1)}\right)$ and $\mathcal{I}_{m+1} = \left(y_{(m)}, \delta_2\right)$, the proximity and Γ_1-regions were only dependent on x and τ (but not on c). Due to scale invariance from Theorem 4.1, we

can assume that each of the end intervals is $(0, 1)$. Let $\Gamma_{1,e}(x, \tau)$ be the Γ_1-region corresponding to $N_e(x, \tau)$ in the end interval case.

First we consider $\tau = 1$ and uniform data in the end intervals. Then for x in the right end interval, $N_e(x, 1) = (0, \min(1, 2x))$ for $x \in (0, 1)$ and the Γ_1-region is $\Gamma_{1,e}(x, 1) = (x/2, 1)$.

Theorem 4.8. *Let $D_{[i]}(1, c)$ be the subdigraph of the central similarity PCD based on uniform data in (δ_1, δ_2) where $\delta_1 < \delta_2$ and $\mathcal{Y}_m$ be a set of m distinct $\mathcal{Y}$ points in (δ_1, δ_2). Then for $i \in \{1, m+1\}$ (i.e., in the end intervals), as $n_i \to \infty$, we have $\sqrt{n_i}\,\left[\rho_{[i]}(1, c) - \mu_e(1)\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu_e(1))$, where $\mu_e(1) = 3/4$ and $4\,\nu_e(1) = 1/24$.*

The Proof is provided in Appendix 1. The sharpest rate of convergence in Theorem 4.8 is $K\,\frac{\mu_e(1)}{\sqrt{n_i\,\nu_e(1)^3}} = 36\sqrt{6}\,\frac{K}{\sqrt{n_i}}$ for $i \in \{1, m+1\}$.

Next we consider the more general case of $\tau > 0$ for the end intervals. By Theorem 4.1, we can assume each end interval to be $(0, 1)$. For $\tau \in (0, 1)$ and x in the right end interval, the proximity region is

$$N_e(x, \tau) = \begin{cases} (x\,(1-\tau), x\,(1+\tau)) & \text{if } x \in (0, 1/(1+\tau)), \\ (x\,(1-\tau), 1) & \text{if } x \in (1/(1+\tau), 1), \end{cases} \tag{26}$$

and the Γ_1-region is

$$\Gamma_{1,e}(x, \tau) = \begin{cases} \left(\frac{x}{1+\tau}, \frac{x}{1-\tau}\right) & \text{if } x \in (0, 1-\tau), \\ \left(\frac{x}{1+\tau}, 1\right) & \text{if } x \in (1-\tau, 1). \end{cases} \tag{27}$$

For $\tau \geq 1$ and x in the right end interval, the proximity region is

$$N_e(x, \tau) = \begin{cases} (0, x\,(1+\tau)) & \text{if } x \in (0, 1/(1+\tau)), \\ (0, 1) & \text{if } x \in (1/(1+\tau), 1), \end{cases} \tag{28}$$

and the Γ_1-region is $\Gamma_{1,e}(x, \tau) = (x/(1+\tau), 1)$.

Theorem 4.9. *Let $D_{[i]}(\tau, c)$ be the subdigraph of the central similarity PCD based on uniform data in (δ_1, δ_2) where $\delta_1 < \delta_2$ and $\mathcal{Y}_m$ be a set of m distinct $\mathcal{Y}$ points in (δ_1, δ_2). Then for $i \in \{1, m+1\}$ (i.e., in the end intervals), and $\tau \in (0, \infty)$, we have $\sqrt{n_i}\,\left[\rho_{[i]}(\tau, c) - \mu_e(\tau)\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu_e(\tau))$, as $n_i \to \infty$, where*

$$\mu_e(\tau) = \begin{cases} \frac{\tau\,(\tau+2)}{2\,(\tau+1)} & \textit{if } 0 < \tau < 1, \\ \frac{1+2\,\tau}{2\,(\tau+1)} & \textit{if } \tau \geq 1, \end{cases} \tag{29}$$

and

$$4\,\nu_e(\tau) = \begin{cases} \frac{\tau^2\left(4\,\tau+4-2\,\tau^4-4\,\tau^3-\tau^2\right)}{3\,(\tau+1)^3} & \textit{if } 0 < \tau < 1, \\ \frac{\tau^2}{3\,(\tau+1)^3} & \textit{if } \tau \geq 1. \end{cases} \tag{30}$$

See Appendix 1 for the proof and Figure 5 for the plots of $\mu_e(\tau)$ and $4\,\nu_e(\tau)$. Notice that $\lim_{\tau\to\infty}\nu_e(\tau) = 0$, so the CLT result fails for $\tau = \infty$. Furthermore, $\lim_{\tau\to 0}\nu_e(\tau) = 0$. For $\tau = 1$, we have $\mu_e(\tau = 1) = 3/4$, and $4\,\nu_e(\tau = 1) = 1/24$, hence as $\tau \to 1$, the distribution of $\rho_{[i]}(\tau, c)$ converges to the one in Theorem 4.8 for $i \in \{1, m+1\}$.. The sharpest rate of convergence in Theorem 4.9 is $K\,\frac{\mu_e(\tau)}{\sqrt{n_i\,\nu_e(\tau)^3}}$ (explicit form not presented) for $i \in \{1, m+1\}$ and is minimized at $\tau \approx 0.58$ which is found numerically as before. We also checked the plot of $\mu_e(\tau)/\sqrt{\nu_e(\tau)^3}$ (not presented) and verified that this is where the global minimum is attained.

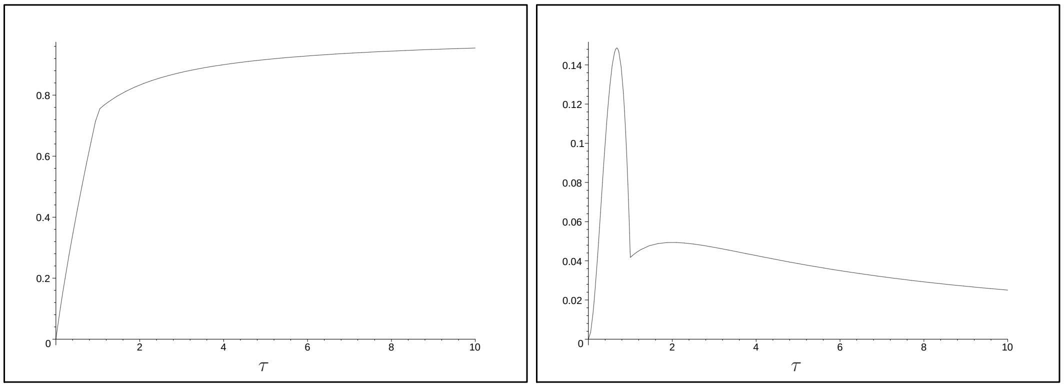

Figure 5: The plots of the asymptotic mean $\mu_e(\tau)$ (left) and the variance $4\,\nu_e(\tau)$ (right) for the end intervals as a function of τ for $\tau \in (0, 10]$.

5 The Distribution of the Relative Density of $\mathcal{U}(\delta_1, \delta_2)$-random $\mathscr{D}_{n,m}(\tau, c)$-digraphs

In this section, we consider the more challenging case of $m > 2$.

5.1 First Version of Relative Density in the Case of $m \geq 2$

Recall that the relative density $\rho_{n,m}(\tau, c)$ is defined as in Equation (8). Letting $w_i = \left(y_{(i+1)} - y_{(i)}\right)/(\delta_2 - \delta_1)$, for $i = 0, 1, 2, \ldots, m$, we obtain the following as a result of Theorem 4.7.

Theorem 5.1. *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(\delta_1, \delta_2)$ with $-\infty < \delta_1 < \delta_2 < \infty$ and $\mathcal{Y}_m$ be a set of m distinct points in (δ_1, δ_2). For $\tau \in (0, \infty)$, the asymptotic distribution of $\rho_{n,m}(\tau, c)$ conditional on $\mathcal{Y}_m$ is given by*

$$\sqrt{n}\left(\rho_{n,m}(\tau, c) - \breve{\mu}(m, \tau, c)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4\,\breve{\nu}(m, \tau, c)\right), \tag{31}$$

as $n \to \infty$, provided that $\breve{\nu}(m, \tau, c) > 0$, where $\breve{\mu}(m, \tau, c) = \widetilde{\mu}(m, \tau, c) \Big/ \left(\sum_{i=1}^{m+1} w_i^2\right)$ with $\widetilde{\mu}(m, \tau, c) = \mu(\tau, c) \sum_{i=2}^{m} w_i^2 + \mu_e(\tau) \sum_{i \in \{1, m+1\}} w_i^2$ and $\mu(\tau, c)$ and $\mu_e(\tau)$ are as in Theorems 4.7 and 4.9, respectively. Furthermore, $4\breve{\nu}(m, \tau, c) = 4\widetilde{\nu}(m, \tau, c) \Big/ \left(\sum_{i=1}^{m+1} w_i^2\right)^2$ with $4\widetilde{\nu}(m, \tau, c) = [P_{2N} + 2\,P_{NG} + P_{2G}] \sum_{i=2}^{m} w_i^3 + [P_{2N,e} + 2\,P_{NG,e} + P_{2G,e}] \sum_{i \in \{1, m+1\}} w_i^3 - (\widetilde{\mu}(m, \tau, c))^2$.

Proof is provided in Appendix 2. Notice that if $y_{(1)} = \delta_1$ and $y_{(m)} = \delta_2$, there are only $m-1$ middle intervals formed by $y_{(i)}$. That is, the end intervals $\mathcal{I}_1 = \mathcal{I}_{m+1} = \emptyset$. Hence in Theorem 5.1, $\breve{\mu}(m, \tau, c) = \mu(\tau, c)$ since $\widetilde{\mu}(m, \tau, c) = \mu(\tau, c) \sum_{i=2}^{m} w_i^2$. Furthermore, $4\breve{\nu}(m, \tau, c) = [P_{2N} + 2\,P_{NG} + P_{2G}] \sum_{i=2}^{m} w_i^3 - (\mu(\tau, c) \sum_{i=2}^{m} w_i^2)^2 = 4\,\nu(m, \tau, c) + \mu^2(\tau, c) \left(\sum_{i=2}^{m} w_i^3 - (\sum_{i=2}^{m} w_i^2)^2\right)$.

5.2 Second Version of Relative Density in the Case of $m \geq 2$

For $m \geq 2$, if we consider the entire data set $\mathcal{X}_n$, then we have n vertices. So we can also consider the relative density as $\widetilde{\rho}_{n,m}(\tau, c) = |\mathcal{A}| / (n\,(n-1))$.

Theorem 5.2. *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(\delta_1, \delta_2)$ with $-\infty < \delta_1 < \delta_2 < \infty$ and $\mathcal{Y}_m$ be a set of m distinct points in (δ_1, δ_2). For $\tau \in (0, \infty)$, the asymptotic distribution for $\widetilde{\rho}_{n,m}(\tau, c)$ conditional on $\mathcal{Y}_m$ is given by*

$$\sqrt{n}\left(\widetilde{\rho}_{n,m}(\tau, c) - \widetilde{\mu}(m, \tau, c)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4\,\widetilde{\nu}(m, \tau, c)\right), \tag{32}$$

as $n \to \infty$, provided that $\widetilde{\nu}(m, \tau, c) > 0$, where $\widetilde{\mu}(m, \tau, c)$ and $\widetilde{\nu}(m, \tau, c)$ are as in Theorem 5.1.

Proof is provided in Appendix 2. Notice that the relative arc densities, $\rho_{n,m}(\tau, c)$ and $\widetilde{\rho}_{n,m}(\tau, c)$ do not have the same distribution for neither finite nor infinite n. But we have $\rho_{n,m}(\tau, c) = \frac{n(n-1)}{n_T}\widetilde{\rho}_{n,m}(\tau, c)$ and since for large n_i and n, $\sum_{i=1}^{m+1} \frac{n_i(n_i-1)}{n(n-1)} \approx \sum_{i=1}^{m+1} w_i^2 < 1$, it follows that $\widetilde{\mu}(m, \tau, c) < \breve{\mu}(m, \tau, c)$ and $\widetilde{\nu}(m, \tau, c) < \breve{\nu}(m, \tau, c)$ for large n_i and n. Furthermore, the asymptotic normality holds for $\rho_{n,m}(\tau, c)$ iff it holds for $\widetilde{\rho}_{n,m}(\tau, c)$.

6 Extension of Central Similarity Proximity Regions to Higher Dimensions

Note that in $\mathbb{R}$ the central similarity PCDs are based on the intervals whose end points are from class $\mathcal{Y}$. This interval partitioning can be viewed as the *Delaunay tessellation* of $\mathbb{R}$ based on $\mathcal{Y}_m$. So in higher dimensions, we use the Delaunay tessellation based on $\mathcal{Y}_m$ to partition the space.

Let $\mathcal{Y}_m = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_m\}$ be m points in general position in $\mathbb{R}^d$ and T_i be the i^{th} Delaunay cell for $i = 1, 2, \ldots, J_m$, where J_m is the number of Delaunay cells. Let $\mathcal{X}_n$ be a set of iid random variables from distribution F in $\mathbb{R}^d$ with support $\mathcal{S}(F) \subseteq \mathcal{C}_H(\mathcal{Y}_m)$ where $\mathcal{C}_H(\mathcal{Y}_m)$ stands for the convex hull of $\mathcal{Y}_m$.

6.1 Extension of Central Similarity Proximity Regions to $\mathbb{R}^2$

For illustrative purposes, we focus on $\mathbb{R}^2$ where a Delaunay tessellation is a *triangulation*, provided that no more than three points in $\mathcal{Y}_m$ are cocircular (i.e., lie on the same circle). Furthermore, for simplicity, we only consider the one Delaunay triangle case. Let $\mathcal{Y}_3 = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\}$ be three non-collinear points in $\mathbb{R}^2$ and $T(\mathcal{Y}_3) = T(\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3)$ be the triangle with vertices $\mathcal{Y}_3$. Let $\mathcal{X}_n$ be a set of iid random variables from F with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$.

For the expansion parameter $\tau \in (0, \infty]$, define $N(x, \tau, M_C)$ to be the *central similarity proximity map* with expansion parameter τ as follows; see also Figure 6. Let e_j be the edge opposite vertex y_j for $j = 1, 2, 3$, and let "edge regions" $R_E(e_1)$, $R_E(e_2)$, $R_E(e_3)$ partition $T(\mathcal{Y}_3)$ using line segments from the center of mass of $T(\mathcal{Y}_3)$ to the vertices. For $x \in (T(\mathcal{Y}_3))^o$, let $e(x)$ be the edge in whose region x falls; $x \in R_E(e(x))$. If x falls on the boundary of two edge regions we assign $e(x)$ arbitrarily. For $\tau > 0$, the central similarity proximity region $N(x, \tau, M_C)$ is defined to be the triangle $T_{CS}(x, \tau) \cap T(\mathcal{Y}_3)$ with the following properties:

(i) For $\tau \in (0, 1]$, the triangle $T_{CS}(x, \tau)$ has an edge $e_\tau(x)$ parallel to $e(x)$ such that $d(x, e_\tau(x)) = \tau\, d(x, e(x))$ and $d(e_\tau(x), e(x)) \leq d(x, e(x))$ and for $\tau > 1$, $d(e_\tau(x), e(x)) < d(x, e_\tau(x))$ where $d(x, e(x))$ is the Euclidean distance from x to $e(x)$,

(ii) the triangle $T_{CS}(x, \tau)$ has the same orientation as and is similar to $T(\mathcal{Y}_3)$,

(iii) the point x is at the center of mass of $T_{CS}(x, \tau)$.

Note that (i) implies the expansion parameter τ, (ii) implies "similarity", and (iii) implies "central" in the name, (parameterized) *central similarity proximity map*. Notice that $\tau > 0$ implies that $x \in N(x, \tau, M_C)$ and, by construction, we have $N(x, \tau, M_C) \subseteq T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3)$. For $x \in \partial(T(\mathcal{Y}_3))$ and $\tau \in (0, \infty]$, we define $N(x, \tau, M_C) = \{x\}$. For all $x \in T(\mathcal{Y}_3)^o$ the edges $e_\tau(x)$ and $e(x)$ are coincident iff $\tau = 1$. Note also that $\lim_{\tau \to \infty} N(x, \tau, M_C) = T(\mathcal{Y}_3)$ for all $x \in (T(\mathcal{Y}_3))^o$, so we define $N(x, \infty, M_C) = T(\mathcal{Y}_3)$ for all such x.

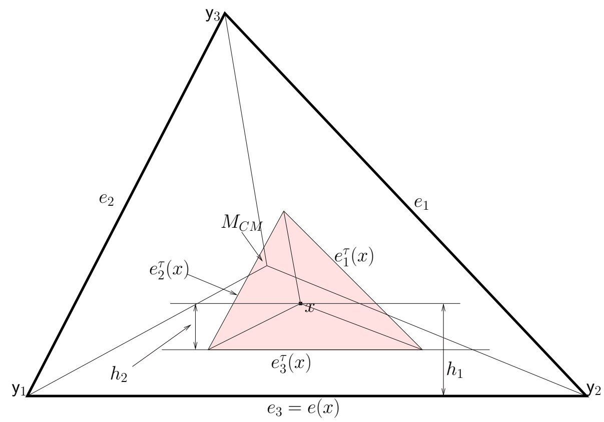

Figure 6: Construction of central similarity proximity region, $N(x, \tau = 1/2, M_C)$ (shaded region) for an $x \in R_E(e_3)$ where $h_2 = d(x, e_3^\tau(x)) = \frac{1}{2}\, d(x, e(x))$ and $h_1 = d(x, e(x))$..

6.2 Extension of Central Similarity Proximity Regions to $\mathbb{R}^d$ with $d > 2$

The extension to $\mathbb{R}^d$ for $d > 2$ with $M = M_C$ is provided in (Ceyhan and Priebe (2005)), the extension for general M is similar: Let $\mathcal{Y} = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_{d+1}\}$ be $d+1$ non-coplanar points. Denote the simplex formed by these $d+1$ points as $\mathfrak{S}(\mathcal{Y}_{d+1})$. The extension of N_{CS}^τ to $\mathbb{R}^d$ for $d > 2$ is straightforward. Let $\mathcal{Y} = \{\mathsf{y}_1, \mathsf{y}_2, \cdots, \mathsf{y}_{d+1}\}$ be $d+1$ points in general position. Denote the simplex formed by these $d+1$ points as $\mathfrak{S}(\mathcal{Y}_{d+1})$. (A simplex is the simplest polytope in $\mathbb{R}^d$ having $d+1$ vertices, $d\,(d+1)/2$ edges and $d+1$ faces of dimension $(d-1)$.) For $\tau \in [0, 1]$, define the central similarity proximity map as follows. Let φ_j be the face opposite vertex y_j for $j = 1, 2, \ldots, d+1$, and "face regions" $R(\varphi_1), \ldots, R(\varphi_{d+1})$ partition $\mathfrak{S}(\mathcal{Y}_{d+1})$ into $d+1$ regions, namely the $d+1$ polytopes with vertices being the center of mass together with d vertices chosen from $d+1$ vertices. For $x \in \mathfrak{S}(\mathcal{Y}_{d+1}) \setminus \mathcal{Y}$, let $\varphi(x)$ be the face in whose region x falls; $x \in R(\varphi(x))$. (If x falls on the boundary of two face regions, we assign $\varphi(x)$ arbitrarily.) For $\tau \in (0, 1]$, the τ-*factor* central similarity proximity region $N(x, \tau, M_C) = N^\tau(x)$ is defined to be the simplex $\mathfrak{S}_\tau(x)$ with the following properties:

(i) $\mathfrak{S}_\tau(x)$ has a face $\varphi_\tau(x)$ parallel to $\varphi(x)$ such that $\tau\, d(x, \varphi(x)) = d(\varphi_\tau(x), x)$ where $d(x, \varphi(x))$ is the Euclidean (perpendicular) distance from x to $\varphi(x)$,

(ii) $\mathfrak{S}_\tau(x)$ has the same orientation as and is similar to $\mathfrak{S}(\mathcal{Y}_{d+1})$,

(iii) x is at the center of mass of $\mathfrak{S}_\tau(x)$. Note that $\tau > 1$ implies that $x \in N(x, \tau, M_C)$.

For $\tau = 0$, define $N(x, \tau, M_C) = \{x\}$ for all $x \in \mathfrak{S}(\mathcal{Y}_{d+1})$.

Theorem 4.1 generalizes, so that any simplex $\mathfrak{S}$ in $\mathbb{R}^d$ can be transformed into a regular polytope (with edges being equal in length and faces being equal in volume) preserving uniformity. Delaunay triangulation becomes Delaunay tessellation in $\mathbb{R}^d$, provided no more than $d+1$ points being cospherical (lying on the boundary of the same sphere). In particular, with $d = 3$, the general simplex is a tetrahedron (4 vertices, 4 triangular faces and 6 edges), which can be mapped into a regular tetrahedron (4 faces are equilateral triangles) with vertices $(0, 0, 0)\,(1, 0, 0)\,(1/2, \sqrt{3}/2, 0)$, $(1/2, \sqrt{3}/6, \sqrt{6}/3)$.

Asymptotic normality of the U-statistic and consistency of the tests hold for $d > 2$.

7 Discussion

In this article, we consider the relative density of a random digraph family called central similarity proximity catch digraph (PCD) which is based on two classes of points (in $\mathbb{R}$). The central similarity PCDs have an expansion parameter $\tau > 0$ and a centrality parameter $c \in (0, 1/2)$. We demonstrate that the relative density of the central similarity PCDs is a U-statistic. Then, applying the central limit theory of the U-statistics, we derive the (asymptotic normal) distribution of the relative density for uniform data for the entire ranges of τ and c. We also determine the parameters τ and c for which the rate of convergence to normality is the fastest.

We can apply the relative density in testing one dimensional bivariate spatial point patterns, as done in Ceyhan et al. (2007) for two-dimensional data. Let $\mathcal{X}$ and $\mathcal{Y}$ be two classes of points which lie in a compact interval in $\mathbb{R}$. Then our null hypothesis is some form of complete spatial randomness of $\mathcal{X}$ points, which implies that distribution of $\mathcal{X}$ points has a uniform distribution in the support interval irrespective of the distribution of the $\mathcal{Y}$ points. The alternatives are the segregation of $\mathcal{X}$ from $\mathcal{Y}$ points or association of $\mathcal{X}$ points with $\mathcal{Y}$ points. In general, association is the pattern in which the points from the two different classes occur close to each other, while segregation is the pattern in which the points from the same class tend to cluster together. In this context, under association, $\mathcal{X}$ points are clustered around $\mathcal{Y}$ points, while under segregation, $\mathcal{X}$ points are clustered away from the $\mathcal{Y}$ points. Notice that we can use the asymptotic distribution (i.e., the normal approximation) of the relative density for spatial pattern tests, so our methodology requires number of $\mathcal{X}$ points to be much larger compared to the number of $\mathcal{Y}$ points. Our results will make the power comparisons possible for data from large families of distributions. Moreover, one might determine the optimal (with respect to empirical size and power) parameter values against segregation and association alternatives.

The central similarity PCDs for one dimensional data can be used in classification as outlined in Priebe et al. (2003a), if a high dimensional data set can be projected to one dimensional space with unsubstantial information loss (by some dimension reduction method). In the classification procedure, one might also determine the optimal parameters (with respect to some penalty function) for the best performance. Furthermore, this work forms the foundation of the generalizations and calculations for uniform and non-uniform cases in multiple dimensions. See Section 6 for the details of the extension to higher dimensions. For example, in $\mathbb{R}^2$, the expansion parameter is still τ, but the centrality parameter is $M = (m_1, m_2)$, which is two dimensional. The optimal parameters for testing spatial patterns and classification can also be determined, as in the one dimensional case.

Acknowledgments

This work was supported by TUBITAK Kariyer Project Grant 107T647.

References

Callaert, H. and Janssen, P. (1978). The Berry-Esseen theorem for U-statistics. *Annals of Statistics*, 6:417–421.

Cannon, A. and Cowen, L. (2000). Approximation algorithms for the class cover problem. In *Proceedings of the 6th International Symposium on Artificial Intelligence and Mathematics*.

Ceyhan, E. (2011a). Relative arc density of an interval catch digraph family. *To appear in Metrika*.

Ceyhan, E. (2011b). Spatial clustering tests based on domination number of a new random digraph family. *Communications in Statistics - Theory and Methods*, 40:1–33.

Ceyhan, E. and Priebe, C. (2003). Central similarity proximity maps in Delaunay tessellations. In *Proceedings of the Joint Statistical Meeting, Statistical Computing Section, American Statistical Association*.

Ceyhan, E. and Priebe, C. E. (2005). The use of domination number of a random proximity catch digraph for testing spatial patterns of segregation and association. *Statistics & Probability Letters*, 73:37–50.

Ceyhan, E. and Priebe, C. E. (2007). On the distribution of the domination number of a new family of parametrized random digraphs. *Model Assisted Statistics and Applications*, 1(4):231–255.

Ceyhan, E., Priebe, C. E., and Marchette, D. J. (2007). A new family of random graphs for testing spatial segregation. *Canadian Journal of Statistics*, 35(1):27–50.

Chartrand, G. and Lesniak, L. (1996). *Graphs & Digraphs*. Chapman & Hall/CRC Press LLC, Florida.

DeVinney, J. and Priebe, C. E. (2006). A new family of proximity graphs: Class cover catch digraphs. *Discrete Applied Mathematics*, 154(14):1975–1982.

DeVinney, J., Priebe, C. E., Marchette, D. J., and Socolinsky, D. (2002). Random walks and catch digraphs in classification. `http://www.galaxy.gmu.edu/interface/I02/I2002Proceedings/DeVinneyJason/DeVinneyJason.paper.pdf`. Proceedings of the 34th Symposium on the Interface: Computing Science and Statistics, Vol. 34.

Erdős, P. and Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae (Debrecen)*, 6:290297.

Janson, S., Łuczak, T., and Ruciński, A. (2000). *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., New York.

Jaromczyk, J. W. and Toussaint, G. T. (1992). Relative neighborhood graphs and their relatives. *Proceedings of IEEE*, 80:1502–1517.

Lehmann, E. L. (1988). *Nonparametrics: Statistical Methods Based on Ranks*. Prentice-Hall, Upper Saddle River, NJ.

Marchette, D. J. and Priebe, C. E. (2003). Characterizing the scale dimension of a high dimensional classification problem. *Pattern Recognition*, 36(1):45–60.

Priebe, C. E., DeVinney, J. G., and Marchette, D. J. (2001). On the distribution of the domination number of random class cover catch digraphs. *Statistics & Probability Letters*, 55:239–246.

Priebe, C. E., Marchette, D. J., DeVinney, J., and Socolinsky, D. (2003a). Classification using class cover catch digraphs. *Journal of Classification*, 20(1):3–23.

Priebe, C. E., Solka, J. L., Marchette, D. J., and Clark, B. T. (2003b). Class cover catch digraphs for latent class discovery in gene expression monitoring by DNA microarrays. *Computational Statistics & Data Analysis on Visualization*, 43-4:621–632.

Prisner, E. (1994). Algorithms for interval catch digraphs. *Discrete Applied Mathematics*, 51:147–157.

Sen, M., Das, S., Roy, A., and West, D. (1989). Interval digraphs: An analogue of interval graphs. *Journal of Graph Theory*, 13:189–202.

Toussaint, G. T. (1980). The relative neighborhood graph of a finite planar set. *Pattern Recognition*, 12(4):261–268.

Tuza, Z. (1994). Inequalities for minimal covering sets in sets in set systems of given rank. *Discrete Applied Mathematics*, 51:187–195.

APPENDIX 1: Proofs for the One Interval Case

Proof of Theorem 4.5:

Depending on the location of x_1, the following are the different types of the combinations of $N(x_1, 1, c)$ and $\Gamma_1(x_1, 1, c)$.

(i) for $0 < x_1 \leq c$, we have $N(x_1, 1, c) = (0, x_1/c)$ and $\Gamma_1(x_1, 1, c) = (c\, x_1, (1-c)\, x_1 + c)$,

(ii) for $c < x_1 < 1$, $N(x_1, 1, c) = ((x_1 - c)/(1-c), 1)$ and $\Gamma_1(x_1, 1, c) = (c\, x_1, (1-c)\, x_1 + c)$.

Then $\mu(1, c) = P(X_2 \in N(X_1, 1, c)) = \int_0^c \frac{x_1}{c} dx_1 + \int_c^1 (1 - \frac{x_1 - c}{1-c}) dx_1 = 1/2$.

For $\mathbf{Cov}(h_{12}, h_{13})$, we need to calculate P_{2N}, P_{NG}, and P_{2G}.

$$P_{2N} = P(\{X_2, X_3\} \subset N(X_1, 1, c)) = \int_0^c \left(\frac{x_1}{c}\right)^2 dx_1 + \int_c^1 \left(1 - \frac{x_1 - c}{1-c}\right)^2 dx_1 = 1/3.$$

$$P_{NG} = P(X_2 \in N(X_1, 1, c), X_3 \in \Gamma_1(X_1, 1, c)) =$$
$$\int_0^c \frac{x_1}{c}(1 + c - 2\, c\, x_1) dx_1 + \int_c^1 \left(1 - \frac{x_1 - c}{1-c}\right)(1 + c - 2\, c\, x_1) dx_1 = -c^2/3 + c/3 + 1/6.$$

Finally, $P_{2G} = P(\{X_2, X_3\} \subset \Gamma_1(X_1, 1, c)) = \int_0^1 (1 + c - 2\, c\, x_1)^2 dx_1 = c^2/3 - c/3 + 1/3$.

Therefore $4\, \mathbf{E}[h_{12} h_{13}] = P_{2N} + 2\, P_{NG} + P_{2G} = -c^2/3 + c/3 + 1$. Hence $4\, \nu(1, c) = 4\, \mathbf{Cov}[h_{12}, h_{13}] = c\,(1-c)/3$. ■

Proof of Theorem 4.6:

There are two cases for τ, namely $0 < \tau < 1$ and $\tau \geq 1$.
Case 1: $0 < \tau < 1$: In this case depending on the location of x_1, the following are the different types of the combinations of $N(x_1, \tau, 1/2)$ and $\Gamma_1(x_1, \tau, 1/2)$.

(i) for $0 < x_1 \leq (1-\tau)/2$, we have $N(x_1, \tau, 1/2) = (x_1\,(1-\tau), x_1\,(1+\tau))$ and $\Gamma_1(x_1, \tau, 1/2) = (x_1/(1+\tau), x_1/(1-\tau))$,

(ii) for $(1-\tau)/2 < x_1 \leq 1/2$, we have $N(x_1, \tau, 1/2) = (x_1\,(1-\tau), x_1\,(1+\tau))$ and $\Gamma_1(x_1, \tau, 1/2) = (x_1/(1+\tau), (x_1+\tau)/(1+\tau))$.

Then $\mu(\tau, 1/2) = P(X_2 \in N(X_1, \tau, 1/2)) = 2\, P(X_2 \in N(X_1, \tau, 1/2), X_1 \in (0, 1/2))$ by symmetry and

$$P(X_2 \in N(X_1, \tau, 1/2), X_1 \in (0, 1/2)) = \int_0^{1/2} (x_1\,(1+\tau) - x_1\,(1-\tau)) dx_1 = \int_0^{1/2} 2\, x_1\, \tau dx_1 = \tau/4.$$

So $\mu(\tau, 1/2) = 2\,(\tau/4) = \tau/2$.

For $\mathbf{Cov}(h_{12}, h_{13})$, we need to calculate P_{2N}, P_{NG}, and P_{2G}.

$$P_{2N} = P(\{X_2, X_3\} \subset N(X_1, \tau, 1/2)) = 2\,P(\{X_2, X_3\} \subset N(X_1, \tau, 1/2), X_1 \in (0, 1/2))$$

and

$$P(\{X_2, X_3\} \subset N(X_1, \tau, 1/2), X_1 \in (0, 1/2)) = \int_0^{1/2} (2\,x_1\,\tau)^2 dx_1 = \tau^2/6.$$

So $P_{2N} = 2\,(\tau^2/6) = \tau^2/3$.

$$P_{NG} = P(X_2 \in N(X_1, \tau, 1/2), X_3 \in \Gamma_1(X_1, \tau, 1/2)) = \\ 2\,P(X_2 \in N(X_1, \tau, 1/2), X_3 \in \Gamma_1(X_1, \tau, 1/2), X_1 \in (0, 1/2))$$

and

$$P(X_2 \in N(X_1, \tau, 1/2), X_3 \in \Gamma_1(X_1, \tau, 1/2), X_1 \in (0, 1/2) = \\ \int_0^{(1-\tau)/2} (2\,x_1\,\tau) \left(\frac{x_1}{1-\tau} - \frac{x_1}{1+\tau}\right) dx_1 + \int_{(1-\tau)/2}^{1/2} (2\,x_1\,\tau) \left(\frac{x_1+\tau}{1+\tau} - \frac{x_1}{1+\tau}\right) dx_1 = \\ \int_0^{(1-\tau)/2} (2\,x_1\,\tau) \left(\frac{2\,x_1\,\tau}{1-\tau^2}\right) dx_1 + \int_{(1-\tau)/2}^{1/2} (2\,x_1\,\tau) \left(\frac{\tau}{1+\tau}\right) dx_1 = \frac{\left(2+2\,\tau-\tau^2\right)\tau^2}{12\,(\tau+1)}.$$

So $P_{NG} = \frac{\left(2+2\,\tau-\tau^2\right)\tau^2}{6\,(\tau+1)}$.

Finally,

$$P_{2G} = P(\{X_2, X_3\} \subset \Gamma_1(X_1, \tau, 1/2)) = 2\,P(\{X_2, X_3\} \subset \Gamma_1(X_1, \tau, 1/2), X_1 \in (0, 1/2))$$

and

$$P(\{X_2, X_3\} \subset \Gamma_1(X_1, \tau, 1/2), X_1 \in (0, 1/2)) = \int_0^{(1-\tau)/2} \left(\frac{2\,x_1\,\tau}{1-\tau^2}\right)^2 dx_1 + \int_{(1-\tau)/2}^{1/2} \left(\frac{\tau}{1+\tau}\right)^2 dx_1 = \frac{\tau^2\,(2\,\tau+1)}{6\,(\tau+1)^2}.$$

So $P_{2G} = \frac{\tau^2(2\,\tau+1)}{6\,(\tau+1)^2}$.

Therefore $4\,\mathbf{E}[h_{12}h_{13}] = P_{2N} + 2\,P_{NG} + P_{2G} = \frac{\tau^2\left(8\,\tau+4-\tau^3+2\,\tau^2\right)}{3\,(\tau+1)^2}$. Hence $4\,\nu(\tau, 1/2) = 4\,\mathbf{Cov}[h_{12}, h_{13}] = \frac{\tau^2\left(-\tau^3-\tau^2+2\,\tau+1\right)}{3\,(\tau+1)^2}$.

Case 2: $\tau \geq 1$: In this case depending on the location of x_1, the following are the different types of the combinations of $N(x_1, \tau, 1/2)$ and $\Gamma_1(x_1, \tau, 1/2)$.

(i) for $0 < x_1 \leq 1/(1+\tau)$, we have $N(x_1, \tau, 1/2) = (0, x_1\,(1+\tau))$ and $\Gamma_1(x_1, \tau, 1/2) = (x_1/(1+\tau), (x_1+\tau)/(1+\tau))$,

(ii) for $1/(1+\tau) < x_1 \leq 1/2$, we have $N(x_1, \tau, 1/2) = (0, 1)$ and $\Gamma_1(x_1, \tau, 1/2) = (x_1/(1+\tau), (x_1+\tau)/(1+\tau))$,

Then $\mu(\tau, 1/2) = P(X_2 \in N(X_1, \tau, 1/2)) = 2\,P(X_2 \in N(X_1, \tau, 1/2), X_1 \in (0, 1/2))$ by symmetry and

$$P(X_2 \in N(X_1, \tau, 1/2), X_1 \in (0, 1/2)) = \int_0^{1/(1+\tau)} x_1\,(1+\tau) dx_1 + \int_{1/(1+\tau)}^{1/2} 1 dx_1 = \frac{\tau}{2\,(\tau+1)}.$$

So $\mu(\tau,1/2) = 2\left(\frac{\tau}{2\,(\tau+1)}\right) = \frac{\tau}{(\tau+1)}$.

Next

$$P_{2N} = P(\{X_2,X_3\} \subset N(X_1,\tau,1/2)) = 2\,P(\{X_2,X_3\} \subset N(X_1,\tau,1/2), X_1 \in (0,1/2))$$

and

$$P(\{X_2,X_3\} \subset N(X_1,\tau,1/2), X_1 \in (0,1/2)) = \int_0^{1/(1+\tau)} (x_1\,(1+\tau))^2 dx_1 + \int_{1/(1+\tau)}^{1/2} 1 dx_1 = \frac{1-3\,\tau}{6\,(\tau+1)}.$$

So $P_{2N} = 2\left(\frac{1-3\,\tau}{6\,(\tau+1)}\right) = \frac{1-3\,\tau}{3\,(\tau+1)}$.

$$P_{NG} = P(X_2 \in N(X_1,\tau,1/2), X_3 \in \Gamma_1(X_1,\tau,1/2)) = \\ 2\,P(X_2 \in N(X_1,\tau,1/2), X_3 \in \Gamma_1(X_1,\tau,1/2), X_1 \in (0,1/2))$$

and

$$P(X_2 \in N(X_1,\tau,1/2), X_3 \in \Gamma_1(X_1,\tau,1/2), X_1 \in (0,1/2) = \\ \int_0^{1/(1+\tau)} (x_1\,(1+\tau))(\tau/(1+\tau))dx_1 + \int_{1/(1+\tau)}^{1/2} (\tau/(1+\tau))dx_1 = \frac{\tau^2}{2\,(1+\tau)^2}.$$

So $P_{NG} = \frac{\tau^2}{(1+\tau)^2}$.

Finally,

$$P_{2G} = P(\{X_2,X_3\} \subset \Gamma_1(X_1,\tau,1/2)) = 2\,P(\{X_2,X_3\} \subset \Gamma_1(X_1,\tau,1/2), X_1 \in (0,1/2))$$

and

$$P(\{X_2,X_3\} \subset \Gamma_1(X_1,\tau,1/2), X_1 \in (0,1/2)) = \int_0^{1/2} (\tau/(1+\tau))^2 dx_1 = \frac{\tau^2}{2\,(1+\tau)^2}.$$

So $P_{2G} = \frac{\tau^2}{(1+\tau)^2}$.

Therefore $4\,\mathbf{E}[h_{12}h_{13}] = P_{2N} + 2\,P_{NG} + P_{2G} = \frac{12\,\tau^2+2\,\tau-1}{3\,(\tau+1)^2}$. Hence $4\,\nu(\tau,1/2) = 4\,\mathbf{Cov}[h_{12},h_{13}] = \frac{2\,\tau-1}{3\,(\tau+1)^2}$.
■

Proof of Theorem 4.7:

First we consider $0 < c \le 1/2$. There are two cases for τ, namely $0 < \tau < 1$ and $\tau \ge 1$.
Case 1: $0 < \tau < 1$: In this case depending on the location of x_1, the following are the different types of the combinations of $N(x_1,\tau,c)$ and $\Gamma_1(x_1,\tau,c)$. Let $a_1 := x_1\,(1-\tau)$, $a_2 := x_1\,(1+\frac{(1-c)\,\tau}{c})$, $a_3 := x_1 - \frac{c\,\tau\,(1-x_1)}{1-c}$, $a_4 := x_1 + (1-x_1)\,\tau$, and $g_1 := \frac{c\,x_1}{c+(1-c)\,\tau}$, $g_2 := \frac{x_1}{1-\tau}$, $g_3 := \frac{x_1-\tau}{1-\tau}$, $g_4 := \frac{x_1\,(1-c)+c\,\tau}{1-c+c\,\tau}$. Then

(i) for $0 < x_1 \le c\,(1-\tau)$, we have $N(x_1,\tau,c) = (a_1,a_2)$ and $\Gamma_1(x_1,\tau,c) = (g_1,g_2)$,

(ii) for $c\,(1-\tau) < x_1 \le c$, we have $N(x_1,\tau,c) = (a_1,a_2)$ and $\Gamma_1(x_1,\tau,c) = (g_1,g_4)$,

(iii) for $c < x_1 \le c\,(1-\tau)+\tau$, we have $N(x_1,\tau,c) = (a_3,a_4)$ and $\Gamma_1(x_1,\tau,c) = (g_1,g_4)$,

(iv) for $c\,(1-\tau)+\tau < x_1 < 1$, we have $N(x_1,\tau,c)=(a_3,a_4)$ and $\Gamma_1(x_1,\tau,c)=(g_3,g_4)$.

Then $\mu(\tau,c)=P(X_2\in N(X_1,\tau,c))=\int_0^c(a_2-a_1)dx_1+\int_c^1(a_4-a_3)dx_1=\tau/2$.

For $\mathbf{Cov}(h_{12},h_{13})$, we need to calculate P_{2N}, P_{NG}, and P_{2G}.

$$P_{2N}=P(\{X_2,X_3\}\subset N(X_1,\tau,c))=\int_0^c(a_2-a_1)^2dx_1+\int_c^1(a_4-a_3)^2dx_1=\tau^2/3.$$

$$P_{NG}=P(X_2\in N(X_1,\tau,c),X_3\in\Gamma_1(X_1,\tau,c))=\int_0^{c\,(1-\tau)}(a_2-a_1)(g_2-g_1)dx_1+$$
$$\int_{c\,(1-\tau)}^{c}(a_2-a_1)(g_4-g_1)dx_1+\int_c^{c\,(1-\tau)+\tau}(a_4-a_3)(g_4-g_1)dx_1+\int_{c\,(1-\tau)+\tau}^{1}(a_4-a_3)(g_4-g_3)dx_1=$$
$$\frac{\tau^2\left(c^2\,\tau^3-5\,c^2\,\tau^2-c\,\tau^3+4\,c^2\,\tau+5\,c\,\tau^2-2\,c^2-4\,c\,\tau-\tau^2+2\,c+2\,\tau\right)}{6\,(c\,\tau-c+1)\,(c+\tau-c\,\tau)}.$$

Finally,

$$P_{2G}=P(\{X_2,X_3\}\subset\Gamma_1(X_1,\tau,c))=\int_0^{c\,(1-\tau)}(g_2-g_1)^2dx_1+$$
$$\int_{c\,(1-\tau)}^{c\,(1-\tau)+\tau}(g_4-g_1)^2dx_1+\int_{c\,(1-\tau)+\tau}^{1}(g_4-g_3)^2dx_1=\frac{\left(2\,c^2\,\tau-c^2-2\,c\,\tau+c+\tau\right)\tau^2}{3\,(c\,\tau-c+1)\,(c+\tau-c\,\tau)}.$$

Therefore

$$4\,\mathbf{E}[h_{12}h_{13}]=P_{2N}+2\,P_{NG}+P_{2G}=\frac{\tau^2\left(c^2\,\tau^3-6\,c^2\,\tau^2-c\,\tau^3+8\,c^2\,\tau+6\,c\,\tau^2-4\,c^2-8\,c\,\tau-\tau^2+4\,c+4\,\tau\right)}{3\,(c\,\tau-c+1)\,(c+\tau-c\,\tau)}.$$

Hence $4\,\kappa_1(\tau,c)=4\,\mathbf{Cov}[h_{12},h_{13}]=\frac{\tau^2\left(c^2\,\tau^3-3\,c^2\,\tau^2-c\,\tau^3+2\,c^2\,\tau+3\,c\,\tau^2-c^2-2\,c\,\tau-\tau^2+c+\tau\right)}{3\,(c\,\tau-c+1)(c+\tau-c\,\tau)}$.

Case 2: $\tau\ge 1$: In this case depending on the location of x_1, the following are the different types of the combinations of $N(x_1,\tau,c)$ and $\Gamma_1(x_1,\tau,c)$.

(i) for $0<x_1\le\frac{c}{c+(1-c)\,\tau}$, we have $N(x_1,\tau,c)=(0,a_2)$ and $\Gamma_1(x_1,\tau,c)=(g_1,g_4)$,

(ii) for $\frac{c}{c+(1-c)\,\tau}<x_1\le\frac{c\,\tau}{1-c+c\,\tau}$, we have $N(x_1,\tau,c)=(0,1)$ and $\Gamma_1(x_1,\tau,c)=(g_1,g_4)$,

(iii) for $\frac{c\,\tau}{1-c+c\,\tau}<x_1<1$, we have $N(x_1,\tau,c)=(a_3,1)$ and $\Gamma_1(x_1,\tau,c)=(g_1,g_4)$.

Then

$$\mu(\tau,c)=P(X_2\in N(X_1,\tau,c))=\int_0^{\frac{c}{c+(1-c)\,\tau}}a_2dx_1+\int_{\frac{c}{c+(1-c)\,\tau}}^{\frac{c\,\tau}{1-c+c\,\tau}}1dx_1+\int_{\frac{c\,\tau}{1-c+c\,\tau}}^{1}(1-a_3)dx_1=$$
$$\frac{\tau\,\left(2\,c^2\,\tau-2\,c^2-2\,c\,\tau+2\,c-1\right)}{2\,(c\,\tau-c+1)\,(c\,\tau-c-\tau)}.$$

Next

$$P_{2N} = P(\{X_2, X_3\} \subset N(X_1, \tau, c)) = \int_0^{\frac{c}{c+(1-c)\tau}} a_2^2 dx_1 + \int_{\frac{c}{c+(1-c)\tau}}^{\frac{c\tau}{1-c+c\tau}} 1 dx_1 + \int_{\frac{c\tau}{1-c+c\tau}}^{1} (1-a_3)^2 dx_1 = \frac{3\,c^2\,\tau^2 - 2\,c^2\,\tau - 3\,c\,\tau^2 - c^2 + 2\,c\,\tau + c - \tau}{3\,(c\,\tau - c + 1)\,(c\,\tau - c - \tau)}.$$

$$P_{NG} = P(X_2 \in N(X_1, \tau, c), X_3 \in \Gamma_1(X_1, \tau, c)) = \int_0^{\frac{c}{c+(1-c)\tau}} a_2\,(g_4 - g_1) dx_1 + \int_{\frac{c}{c+(1-c)\tau}}^{\frac{c\tau}{1-c+c\tau}} (g_4 - g_1) dx_1 + \int_{\frac{c\tau}{1-c+c\tau}}^{1} (1-a_3)\,(g_4 - g_1) dx_1 =$$

$$[\tau^2\big(6\,c^6\,\tau^4 - 24\,c^6\,\tau^3 - 18\,c^5\,\tau^4 + 36\,c^6\,\tau^2 + 72\,c^5\,\tau^3 + 18\,c^4\,\tau^4 - 24\,c^6\,\tau - 108\,c^5\,\tau^2 - 84\,c^4\,\tau^3 - 6\,c^3\,\tau^4 + 6\,c^6 + 72\,c^5\,\tau + 132\,c^4\,\tau^2 + 48\,c^3\,\tau^3 - 18\,c^5 - 92\,c^4\,\tau - 84\,c^3\,\tau^2 - 12\,c^2\,\tau^3 + 26\,c^4 + 64\,c^3\,\tau + 30\,c^2\,\tau^2 - 22\,c^3 - 26\,c^2\,\tau - 6\,c\,\tau^2 + 10\,c^2 + 6\,c\,\tau - 2\,c - \tau\big)]/[6\,(c\,\tau - c + 1)^3\,(c\,\tau - c - \tau)^3].$$

Finally,

$$P_{2G} = P(\{X_2, X_3\} \subset \Gamma_1(X_1, \tau, c)) = \int_0^1 (g_4 - g_1)^2 dx_1 = \frac{\big(3\,c^4\,\tau^2 - 6\,c^4\,\tau - 6\,c^3\,\tau^2 + 3\,c^4 + 12\,c^3\,\tau + 3\,c^2\,\tau^2 - 6\,c^3 - 9\,c^2\,\tau + 7\,c^2 + 3\,c\,\tau - 4\,c + 1\big)\,\tau^2}{3\,(c\,\tau - c + 1)^2\,(c\,\tau - c - \tau)^2}.$$

Therefore

$$4\,\mathbf{E}[h_{12}h_{13}] = P_{2N} + 2\,P_{NG} + P_{2G} = [12\,c^6\,\tau^6 - 50\,c^6\,\tau^5 - 36\,c^5\,\tau^6 + 79\,c^6\,\tau^4 + 150\,c^5\,\tau^5 + 36\,c^4\,\tau^6 - 56\,c^6\,\tau^3 - 237\,c^5\,\tau^4 - 175\,c^4\,\tau^5 - 12\,c^3\,\tau^6 + 14\,c^6\,\tau^2 + 168\,c^5\,\tau^3 + 297\,c^4\,\tau^4 + 100\,c^3\,\tau^5 + 2\,c^6\,\tau - 42\,c^5\,\tau^2 - 220\,c^4\,\tau^3 - 199\,c^3\,\tau^4 - 25\,c^2\,\tau^5 - c^6 - 6\,c^5\,\tau + 58\,c^4\,\tau^2 + 160\,c^3\,\tau^3 + 75\,c^2\,\tau^4 + 3\,c^5 + 7\,c^4\,\tau - 46\,c^3\,\tau^2 - 70\,c^2\,\tau^3 - 15\,c\,\tau^4 - 3\,c^4 - 4\,c^3\,\tau + 20\,c^2\,\tau^2 + 18\,c\,\tau^3 + c^3 + c^2\,\tau - 4\,c\,\tau^2 - 3\,\tau^3]/[3\,(c\,\tau - c + 1)^3\,(c\,\tau - c - \tau)^3].$$

Hence

$$4\,\kappa_2(\tau, c) = 4\,\mathbf{Cov}[h_{12}, h_{13}] = [c\,(1-c)\,\big(2\,c^4\,\tau^5 - 7\,c^4\,\tau^4 - 4\,c^3\,\tau^5 + 8\,c^4\,\tau^3 + 14\,c^3\,\tau^4 + 3\,c^2\,\tau^5 - 2\,c^4\,\tau^2 - 16\,c^3\,\tau^3 - 7\,c^2\,\tau^4 - c\,\tau^5 - 2\,c^4\,\tau + 4\,c^3\,\tau^2 + 12\,c^2\,\tau^3 + c^4 + 4\,c^3\,\tau - 6\,c^2\,\tau^2 - 4\,c\,\tau^3 - 2\,c^3 - 3\,c^2\,\tau + 4\,c\,\tau^2 + c^2 + c\,\tau - \tau^2\big)]/[3\,(c\,\tau - c + 1)^3\,(c\,\tau - c - \tau)^3].$$

For $1/2 \le c < 1$, by symmetry, it follows that $\mu_2(\tau, c) = \mu_1(\tau, 1-c)$ and $\nu_2(\tau, c) = \nu_1(\tau, 1-c)$. ■

Proof of Theorem 4.8:

Suppose $i = m+1$ (i.e., the support is the right end interval). For $x_1 \in (0, 1)$, depending on the location of x_1, the following are the different types of the combinations of $N_e(x_1, 1)$ and $\Gamma_{1,e}(x_1, 1)$.

(i) for $0 < x_1 \le 1/2$, we have $N_e(x_1, 1) = (0, 2\,x_1)$ and $\Gamma_{1,e}(x_1, 1) = (x_1/2, 1)$,

(ii) for $1/2 < x_1 < 1$, $N_e(x_1, 1) = (0, 1)$ and $\Gamma_{1,e}(x_1, 1) = (x_1/2, 1)$.

Then $\mu_e(1) = P(X_2 \in N_e(X_1, 1)) = \int_0^{1/2} 2x_1 dx_1 + \int_{1/2}^1 1 dx_1 = 3/4$.

For $\mathbf{Cov}(h_{12}, h_{13})$, we need to calculate P_{2N}, P_{NG}, and P_{2G}.

$$P_{2N} = P(\{X_2, X_3\} \subset N_e(X_1, 1)) = \int_0^{1/2} (2x_1)^2 dx_1 + \int_{1/2}^1 1 dx_1 = 2/3.$$

$$P_{NG} = P(X_2 \in N_e(X_1, 1), X_3 \in \Gamma_{1,e}(X_1, 1)) = \int_0^{1/2} (2x_1)(1 - x_1/2) dx_1 + \int_{1/2}^1 1(1 - x_1/2) dx_1 = 25/48.$$

Finally, $P_{2G} = P(\{X_2, X_3\} \subset \Gamma_{1,e}(X_1, 1)) = \int_0^1 (1 - x_1/2)^2 dx_1 = 7/12$.

Therefore $4\,\mathbf{E}[h_{12}h_{13}] = P_{2N} + 2\,P_{NG} + P_{2G} = 55/24$. Hence $4\,\nu_e(1) = 4\,\mathbf{Cov}[h_{12}, h_{13}] = 1/24$.

For uniform data, by symmetry, the distribution of the relative density of the subdigraph for $i = 1$ is identical to $i = m + 1$ case. $\blacksquare$

Proof of Theorem 4.9:

There are two cases for τ, namely, $0 < \tau < 1$ and $\tau \geq 1$.

Case 1: $0 < \tau < 1$: For $x_1 \in (0, 1)$, depending on the location of x_1, the following are the different types of the combinations of $N_e(x_1, \tau)$ and $\Gamma_{1,e}(x_1, \tau)$.

(i) for $0 < x_1 \leq 1 - \tau$, we have $N_e(x_1, \tau) = (x_1\,(1 - \tau), x_1\,(1 + \tau))$ and $\Gamma_{1,e}(x_1, \tau) = (x_1/(1 + \tau), x_1/(1 - \tau))$,

(ii) for $1 - \tau < x_1 \leq 1/(1 + \tau)$, we have $N_e(x_1, \tau) = (x_1\,(1 - \tau), x_1\,(1 + \tau))$ and $\Gamma_{1,e}(x_1, \tau) = (x_1/(1 + \tau), 1)$,

(iii) for $1/(1 + \tau) < x_1 < 1$, we have $N_e(x_1, \tau) = (x_1\,(1 - \tau), 1)$ and $\Gamma_{1,e}(x_1, \tau) = (x_1/(1 + \tau), 1)$.

Then

$$\mu_e(\tau) = P(X_2 \in N_e(X_1, \tau)) = \int_0^{1/(1+\tau)} (x_1\,(1 + \tau) - x_1\,(1 - \tau)) dx_1 + \int_{1/(1+\tau)}^1 (1 - x_1\,(1 - \tau)) dx_1 =$$
$$\int_0^{1/(1+\tau)} (2\,x_1\,\tau) dx_1 + \int_{1/(1+\tau)}^1 (1 - x_1 + x_1\,\tau) dx_1 = \frac{\tau\,(\tau + 2)}{2\,(\tau + 1)}.$$

For $\mathbf{Cov}(h_{12}, h_{13})$, we need to calculate $P_{2N,e}$, $P_{NG,e}$, and $P_{2G,e}$.

$$P_{2N,e} = P(\{X_2, X_3\} \subset N_e(X_1, \tau)) = \int_0^{1/(1+\tau)} (2\,x_1\,\tau)^2 dx_1 + \int_{1/(1+\tau)}^1 (1 - x_1 + x_1\,\tau)^2 dx_1 = \frac{\tau^2\,(\tau^2 + 3\,\tau + 4)}{3\,(\tau + 1)^2}.$$

$$P_{NG,e} = P(X_2 \in N_e(X_1,\tau), X_3 \in \Gamma_{1,e}(X_1,\tau)) = \int_0^{1-\tau} (2\,x_1\,\tau)\left(\frac{2\,x_1\,\tau}{1-\tau^2}\right) dx_1 +$$
$$\int_{1-\tau}^{1/(1+\tau)} (2\,x_1\,\tau)\left(1-\frac{x_1}{1+\tau}\right) dx_1 + \int_{1/(1+\tau)}^{1} (1-x_1\,(1-\tau))\left(1-\frac{x_1}{1+\tau}\right) dx_1 = \frac{\left(7\,\tau^2+14\,\tau+8-2\,\tau^4-2\,\tau^3\right)\tau^2}{6\,\left(\tau+1\right)^3}.$$

Finally,

$$P_{2G,e} = P(\{X_2, X_3\} \subset \Gamma_{1,e}(X_1,\tau)) = \int_0^{1-\tau} \left(\frac{2\,x_1\,\tau}{1-\tau^2}\right)^2 dx_1 + \int_{1-\tau}^{1} \left(1-\frac{x_1}{1+\tau}\right)^2 dx_1 = \frac{\tau^2\,(3\,\tau+4)}{3\,(\tau+1)^2}.$$

Therefore $4\,\mathbf{E}[h_{12}h_{13}] = P_{2N,e} + 2\,P_{NG,e} + P_{2G,e} = \frac{\tau^2\left(2\,\tau^2+5\,\tau+4\right)\left(2\,\tau+4-\tau^2\right)}{3\,(\tau+1)^3}$. Hence

$$4\,\nu_e(\tau) = 4\,\mathbf{Cov}[h_{12},h_{13}] = \frac{\tau^2\,\left(4\,\tau+4-2\,\tau^4-4\,\tau^3-\tau^2\right)}{3\,\left(\tau+1\right)^3}.$$

Case 2: $\tau \geq 1$: For $x_1 \in (0,1)$, depending on the location of x_1, the following are the different types of the combinations of $N_e(x_1,\tau)$ and $\Gamma_{1,e}(x_1,\tau)$.

(i) for $0 < x_1 \leq 1/(1+\tau)$, we have $N_e(x_1,\tau) = (0, x_1\,(1+\tau))$ and $\Gamma_{1,e}(x_1,\tau) = (x_1/(1+\tau), 1)$,

(ii) for $1/(1+\tau) < x_1 < 1$, we have $N_e(x_1,\tau) = (0,1)$ and $\Gamma_{1,e}(x_1,\tau) = (x_1/(1+\tau), 1)$.

Then

$$\mu_e(\tau) = P(X_2 \in N_e(X_1,\tau)) = \int_0^{1/(1+\tau)} x_1\,(1+\tau) dx_1 + \int_{1/(1+\tau)}^{1} 1 dx_1 = \frac{1+2\,\tau}{2\,(\tau+1)}.$$

Next,

$$P_{2N,e} = P(\{X_2, X_3\} \subset N_e(X_1,\tau)) = \int_0^{1/(1+\tau)} (x_1\,(1+\tau))^2 dx_1 + \int_{1/(1+\tau)}^{1} 1 dx_1 = \frac{1+3\,\tau}{3\,(\tau+1)}.$$

$$P_{NG,e} = P(X_2 \in N_e(X_1,\tau), X_3 \in \Gamma_{1,e}(X_1,\tau)) =$$
$$\int_0^{1/(1+\tau)} (x_1\,(1+\tau))\left(1-\frac{x_1}{1+\tau}\right) dx_1 + \int_{1/(1+\tau)}^{1} \left(1-\frac{x_1}{1+\tau}\right) dx_1 = \frac{6\,\tau^3+12\,\tau^2+6\,\tau+1}{6\,\left(\tau+1\right)^3}.$$

Finally,

$$P_{2G,e} = P(\{X_2, X_3\} \subset \Gamma_{1,e}(X_1,\tau)) = \int_0^{1} \left(1-\frac{x_1}{1+\tau}\right)^2 dx_1 = \frac{3\,\tau^2+3\,\tau+1}{3\,(\tau+1)^2}.$$

Therefore $4\,\mathbf{E}[h_{12}h_{13}] = P_{2N,e} + 2\,P_{NG,e} + P_{2G,e} = \frac{12\,\tau^3+25\,\tau^2+15\,\tau+3}{3\,(\tau+1)^3}$. Hence $4\,\nu_e(\tau) = 4\,\mathbf{Cov}[h_{12},h_{13}] = \frac{\tau^2}{3\,(\tau+1)^3}$. ■

APPENDIX 2: Proofs for the Multiple Interval Case

We give the proof of Theorem 5.2 first.

Proof of Theorem 5.2:

Recall that $\widetilde{\rho}_{n,m}(\tau,c)$ is the relative arc density of the PCD for the $m>2$ case. Then it follows that $\widetilde{\rho}_{n,m}(\tau,c)$ is a U-statistic of degree two, so we can write it as $\widetilde{\rho}_{n,m}(\tau,c)=\frac{2}{n(n-1)}\sum_{i<j}h_{ij}$ where $h_{ij}=(g_{ij}+g_{ji})/2$. Then the expectation of $\widetilde{\rho}_{n,m}(\tau,c)$ is

$$\mathbf{E}\left[\widetilde{\rho}_{n,m}(\tau,c)\right]=\frac{2}{n\,(n-1)}\sum\sum_{i<j}\mathbf{E}\left[h_{ij}\right]=\mathbf{E}\left[h_{12}\right]=\mathbf{E}\left[g_{12}\right]=P((X_1,X_2)\in\mathcal{A})=\widetilde{\mu}(m,\tau,c).$$

But, by definition of $N(\cdot,\tau,c)$, if X_1 and X_2 are in different intervals, then $P((X_1,X_2)\in\mathcal{A})=0$. So, by the law of total probability, we have

$$\begin{aligned}\widetilde{\mu}(m,\tau,c):=P((X_1,X_2)\in\mathcal{A})=&\\ &\sum_{i=1}^{m+1}P((X_1,X_2)\in\mathcal{A}\,|\,\{X_1,X_2\}\subset\mathcal{I}_i)\,P(\{X_1,X_2\}\subset\mathcal{I}_i)=\\ &\sum_{i=2}^{m}\mu(\tau,c)\,P(\{X_1,X_2\}\subset\mathcal{I}_i)+\sum_{i\in\{1,m+1\}}\mu_e(\tau)\,P(\{X_1,X_2\}\subset\mathcal{I}_i)=\\ &\sum_{i=2}^{m}\mu(\tau,c)\,w_i^2+\sum_{i\in\{1,m+1\}}\mu_e(\tau)\,w_i^2=\mu(\tau,c)\sum_{i=2}^{m}w_i^2+\mu_e(\tau)\sum_{i\in\{1,m+1\}}w_i^2.\end{aligned}$$

since $P(X_2\in N(X_1,\tau,c)\,|\,\{X_1,X_2\}\subset\mathcal{I}_i)$ is $\mu(\tau,c)$ for middle intervals and $\mu_e(\tau)$ for the end intervals and $P(\{X_1,X_2\}\subset\mathcal{I}_i)=\left(\frac{y_{(i)}-y_{(i-1)}}{\delta_2-\delta_1}\right)^2=w_i^2$.

Furthermore, the asymptotic variance is

$$4\,\widetilde{\nu}(m,\tau,c)=4\,\mathbf{E}\left[h_{12}h_{13}\right]-\mathbf{E}\left[h_{12}\right]\mathbf{E}\left[h_{13}\right]=4\,\mathbf{E}\left[h_{12}h_{13}\right]-(\widetilde{\mu}(m,\tau,c))^2$$

where $4\,\mathbf{E}\left[h_{12}h_{13}\right]=\widetilde{P}_{2N}+2\widetilde{P}_{NG}+\widetilde{P}_{2G}$ with

$$\begin{aligned}\widetilde{P}_{2N}=&\sum_{i=2}^{m}P(\{X_2,X_3\}\subset N(X_1,\tau,c)\,|\,\{X_1,X_2,X_3\}\subset\mathcal{I}_i)\,P(\{X_1,X_2,X_3\}\subset\mathcal{I}_i)+\\ &\sum_{i\in\{1,m+1\}}P(\{X_2,X_3\}\subset N_e(X_1,\tau)\,|\,\{X_1,X_2,X_3\}\subset\mathcal{I}_i)\,P(\{X_1,X_2,X_3\}\subset\mathcal{I}_i)=\\ &\sum_{i=2}^{m}P_{2N}\,P(\{X_1,X_2,X_3\}\subset\mathcal{I}_i)+\sum_{i\in\{1,m+1\}}P_{2N,e}\,P(\{X_1,X_2,X_3\}\subset\mathcal{I}_i)\approx\\ &\sum_{i=2}^{m}P_{2N}\,w_i^3+\sum_{i\in\{1,m+1\}}P_{2N,e}\,w_i^3=P_{2N}\sum_{i=2}^{m}w_i^3+P_{2N,e}\sum_{i\in\{1,m+1\}}w_i^3.\end{aligned}$$

since $P(\{X_2,X_3\}\subset N(X_1,\tau,c)\,|\,\{X_1,X_2,X_3\}\subset\mathcal{I}_i)$ is P_{2N} for middle intervals and $P_{2N,e}$ for the end intervals and $P(\{X_1,X_2,X_3\}\subset\mathcal{I}_i)=\left(\frac{y_{(i)}-y_{(i-1)}}{\delta_2-\delta_1}\right)^3=w_i^3$. Similarly,

$$\widetilde{P}_{NG}=P_{NG}\sum_{i=2}^{m}w_i^3+P_{NG,e}\sum_{i\in\{1,m+1\}}w_i^3$$

and

$$\widetilde{P}_{2G} = P_{2G} \sum_{i=2}^{m} w_i^3 + P_{2G,e} \sum_{i \in \{1,m+1\}} w_i^3.$$

Therefore,

$$4\,\widetilde{\nu}(m,\tau,c) \;=\; (P_{2N} + 2\,P_{NG} + P_{2G}) \sum_{i=2}^{m} w_i^3 + (P_{2N,e} + 2\,P_{NG,e} + P_{2G,e}) \sum_{i \in \{1,m+1\}} w_i^3 - (\widetilde{\mu}(m,\tau,c))^2.$$

Hence the desired result follows. ■

Proof of Theorem 5.1:

Recall that $\rho_{n,m}(\tau,c)$ is the version I of the relative arc density of the PCD for the $m > 2$ case. Moreover, $\rho_{n,m}(\tau,c) = \frac{n(n-1)}{n_T}\widetilde{\rho}_{n,m}(\tau,c)$. Then the expectation of $\rho_{n,m}(\tau,c)$, for large n_i and n, is

$$\mathbf{E}\left[\rho_{n,m}(\tau,c)\right] = \frac{n(n-1)}{n_T}\mathbf{E}[\widetilde{\rho}_{n,m}(\tau,c)] \approx \widetilde{\mu}(m,\tau,c) \left(\sum_{i=1}^{m+1} w_i^2\right)^{-1}$$

since $\frac{n(n-1)}{n_T} = \left(\sum_{i=1}^{m+1} n_i(n_i-1)/(n(n-1))\right)^{-1} \approx \left(\sum_{i=1}^{m+1} w_i^2\right)^{-1}$ for large n_i and n. Here $\widetilde{\mu}(m,\tau,c)$ is as in Theorem 5.2.

Moreover, the asymptotic variance of $\rho_{n,m}(\tau,c)$, for large n_i and n, is

$$4\,\breve{\nu}(m,\tau,c) = \frac{n^2(n-1)^2}{n_T^2} 4\,\widetilde{\nu}(m,\tau,c) = 4\,\widetilde{\nu}(m,\tau,c) \left(\sum_{i=1}^{m+1} w_i^2\right)^{-2}$$

since

$$\frac{n^2(n-1)^2}{n_T^2} = \left(\sum_{i=1}^{m+1} n_i(n_i-1)/(n(n-1))\right)^{-2} \approx \left(\sum_{i=1}^{m+1} w_i^2\right)^{-2}$$

for large n_i and n, Here $\widetilde{\nu}(m,\tau,c)$ is as in Theorem 5.2. Hence the desired result follows. ■